# Psychometrika

VOLUME XII — 1947
JANUARY-DECEMBER



PUBLISHED QUARTERLY
By THE PSYCHOMETRIC SOCIETY
AT 23 WEST COLORADO AVENUE
COLORADO SPRINGS, COLORADO

# Psychometrika

## CONTENTS



VOLUME TWELVE MARCH 1947 NUMBER ONE

# CHAPTER ONE

PSYCHOMETRIKA—VOL. 12, NO. 1
MARCH, 1947

# TEST "RELIABILITY": ITS MEANING AND DETERMINATION

LEE J. CRONBACH
UNIVERSITY OF CHICAGO

The concept of test reliability is examined in terms of general, group, and specific factors among the items, and the stability of scores in these factors from trial to trial. Four essentially different definitions of reliability are distinguished, which may be called the hypothetical self-correlation, the coefficient of equivalence, the coefficient of stability, and the coefficient of stability and equivalence. The possibility of estimating each of these coefficients is discussed. The coefficients are not interchangeable and have different values in corrections for attentuation, standard errors of measurement, and other practical applications.

The literature of testing contains many discussions of test reliability. Each year, new formulations are offered, and new procedures for estimating reliability are championed. There appears to have developed no universally accepted procedure, and several writers have attributed this difficulty to the diversity of definitions for reliability now in use. It has often been suggested that perhaps the only effective way to resolve the conflicts among contending viewpoints is to replace the term "reliability," recognizing that it covers not one, but several concepts. The present paper attempts to restate the conflicting concepts and assumptions now current, and to offer a scheme for separating the various aspects of dependability of measurement.

The physical scientist generally has expressed the accuracy of his observations in terms of the variation of repeated observations of the same event. The mean of the squared deviations of these observations about the obtained mean is the "error variance." This is a measure of precision or reliability. If for the present we regard reliability as the consistency of repeated measurements of the same event by the same process, two fundamental differences between the problem of the physical scientist and the psychologist appear. The physical scientist makes two assumptions, both of which are adequately true for him. First, he assumes that the entity being measured does not change during the measurement process. By controlling the relevant conditions—and he usually knows what these conditions are and can control them—he can hold nearly constant the length of a rod or the pressure of a gas. When measuring a variable

quantity, where his assumption is no longer valid, he abandons the method of successive observations and employs instead simultaneous observations. The psychologist cannot obtain simultaneous measurements of behavior, yet the quantities that interest him are always variable and the method of successive measurements requires an impossible assumption. The psychologist may wish to measure a hypothetical constant (aptitude, or a limen), but all he can ever observe is behavior, which is always shifting. It is one thing to test the accuracy of measurement of a quantity, quite another to test whether that quantity is constant. Judgment on the second question must await judgment on the first.

The second assumption of the physical scientist is that his measurements are independent. If one rules out his remembering prior measurements, this assumption can usually be made true. Successive measurements of psychological quantities are rarely independent, however, because the act of measurement may change the quantity. London (10) has recently described this difficulty by the physicist's term *hysteresis*.

The reliability of a test score has generally been defined in terms of the variation of scores obtained by the individual on successive independent testings. Neither the assumption of constancy of true scores nor the assumption of experimental independence is realized in practice with most psychological variables; therefore, *the reliability of a test, as so defined, is a concept which cannot be directly observed.* If there is no standard of truth, it is fruitless to compare one estimate with another and debate which is more correct. But by various assumptions which usually cannot be tested, we obtain usable statistics which describe the test. *Different assumptions lead to different types of coefficients, which are not estimates of each other.* In particular, as many writers have noted, an estimate of the stability of a test score is not at all the same as an estimate of the accuracy of measurement of behavior at any one instant. Jenkins cites Franzen's comments on certain physiological measures which have high split-half "reliabilities" and low retest "reliabilities" (6). The measuring technique may be extremely accurate in reporting a biological instant in the life of an individual but not measure a stable characteristic of the individual.

Both the physicist and the psychologist encounter the problem of observer error. In sighting through a telescope or scoring an essay test, there is likely to be appreciable constant and variable error in observing. If one compares several judgments by the same observer, he includes the variable errors of observation with the errors of measurement. Hence, he studies the reliability of "this measuring

instrument used by this man." If scores obtained by several observers in simultaneous measurements are pooled for comparison, the constant error of each man is included as a source of variation. This procedure studies the reliability of "this measuring instrument used by different men." Since the human takes part in the measurement, one cannot study the reliability of an instrument apart from the men who use it.

### *Types of "Reliability"*

It is known that

$$r_{tt} = 1 - \frac{\sigma_e^2}{\sigma_t^2}, \tag{1}$$

where $r_{tt}$ is the reliability coefficient, $\sigma_e^2$ is the hypothetical error variance—the mean of the squared deviations of all obtained scores for each person from the mean obtained score for that person—and $\sigma_t^2$ is the variance of the scores of all persons on all the hypothetical independent trials.

It is convenient to consider the possible definitions of error of measurement in terms of variance. Using a bi-factor pattern to describe a test,* the variance of scores from a single testing may be expressed as follows:

$$\sigma_1^2 = \sigma_g^2 + \sigma_{f_1}^2 + \sigma_{f_2}^2 + \cdots + \sigma_{s_1}^2 + \sigma_{s_2}^2 + \cdots + \sigma_{s_n}^2 + \sigma_{\varepsilon_1}^2. \tag{2}$$

The terms have the following meanings:

$\sigma_1^2$ is the variance of obtained scores;

$\sigma_g^2$ is the variance in the general factor (if any) represented in the test items;

$\sigma_{f_1}^2$, $\sigma_{f_2}^2$, etc., are the respective variances in the orthogonal group factors of undetermined number, each of which is represented in two or more items;

$\sigma_{s_1}^2$, $\sigma_{s_2}^2$, etc., are the respective "specificities" of the $n$ items—the part of the reliable variance of scores on the items which cannot be assigned to common factors; and $\sigma_{\varepsilon_1}^2$ like the residual variance.

The referents for these factors may be illustrated in a hypothetical examination in psychology. The general factor might include general knowledge of psychology, reading ability, motivation, and

* Another factor pattern could be assumed without changing the basic argument (4, 7-9, 107).

other characteristics. Group factors might be related to knowledge of separate topics, mathematical skill required in only a few items, and so on. Each item taps, in addition, some specific knowledge not demanded by other items. The specificity variance accounts for individual differences in these elements. The remaining variance may include momentary inattention, guessing, and other random elements.

For reference, the formula will be rewritten thus:

$$\sigma_1^2 = \sigma_g^2 + \sum \sigma_f^2 + \sum \sigma_{s_i}^2 + \sigma_{\varepsilon_1}^2. \tag{3}$$

Consider now the scores obtained from a series of independent measurements of the same individuals using the same test.

$$\sigma_t^2 = \sigma^2_{\bar{g}_x} + \sum \sigma^2_{\bar{f}_x} + \sum \sigma^2_{\bar{s}_{ix}} + \sum \sigma^2_{g_x} + \sum\sum \sigma^2_{f_x} + \sum\sum \sigma^2_{s_{ix}} + \sigma_{\varepsilon_t}^2. \tag{4}$$

$\sigma_t^2$ is the variance of all obtained scores about the grand mean;

$\sigma^2_{\bar{g}_x}$ is the variance of the mean general factor scores of all individuals about the mean for all individuals—the between-persons variance in $g$;

$\sigma^2_{\bar{f}_x}$ is the between-persons variance in a group factor;

$\sigma^2_{\bar{s}_{ix}}$ is the between-persons variance in specificity on any item;

$\sum \sigma_{g_x}^2$ is the sum over individuals of the variances of the general-factor scores for each individual about the mean for that individual—the within-persons variance;

$\sum\sum \sigma_{f_x}^2$ and $\sum\sum \sigma_{s_{ix}}^2$ represent the corresponding within-persons variances in the group factors and specificities, respectively; and

$\sigma_{\varepsilon_t}^2$ is the residual variance.

The between-persons variances represent, as in the case of the single trial, individual differences in the factors. The within-persons variances represent instability of scores for each individual, as a result of changes from test to test.

These formulations permit an exact statement of what a "reliability coefficient" represents. Apparently at least four fundamentally different meanings of reliability are current:

(1) The "error variance" may be permitted to include, in equation (4), the terms $\sum \sigma_{g_x}^2$, $\sum\sum \sigma_{f_x}^2$, $\sum\sum \sigma_{s_{ix}}^2$, and $\sigma_{\varepsilon_t}^2$. That is, instability is regarded as an error of measurement. This is the coefficient defined by the correlation from repeated independent administrations

of the same test. The assumption of constancy is made, since any change of score from trial to trial is treated as an error of measurement. If that assumption is true, the instability terms vanish, but such constancy in all the behaviors a test measures is highly unlikely.

(2) The "error variance" may be permitted to include, in equation (4), the terms $\sum \sigma^2_{s_{ix}}$, $\sum \sigma^2_{g_x}$, $\sum \sum \sigma^2_{f_x}$, $\sum \sum \sigma^2_{s_{ix}}$, and $\sigma_{\varepsilon_i}^2$. Both instability and specificity are treated as errors. This is the "reliability" defined by the correlation between successive independent administrations of equivalent tests. Because different items are used in preparing equivalent forms, the specific-factor scores of individuals on the two tests will be uncorrelated. These, therefore, contribute to changes in score and are treated as error. If the tests do not represent the same group factors, at least part of $\sum \sigma^2_{f_x}$ is also added to the error variance.

(3) The "error variance" may be permitted to include in equation (3), the terms $\sigma_{s_i}^2$ and $\sigma_{\varepsilon_1}^2$. This defines "reliability" as the correlation between two equivalent tests administered simultaneously. Instability is excluded from consideration, and no assumptions of constancy are made. Specific-factor variances are included in errors of measurement. Depending on the degree of equivalence, part of the group-factor variance may also be treated as error.

(4) The "error variance" may be restricted, in equation (3), to the term $\sigma_{\varepsilon_1}^2$. This is "reliability" defined as the self-correlation of a test (see below). No assumption of constancy is made, and independence is not involved. The specific factors remain the same from test to test and are added to the true-score variance. All real variables measured by the test are treated as quantities estimated, not as errors.

It may now be helpful to restate these definitions and to give them names for reference.

> *Definition* (1): *Reliability is the degree to which the test score indicates unchanging* individual differences in any traits.* (*Coefficient of stability*).
>
> *Definition* (2): *Reliability is the degree to which the test score indicates unchanging individual differences in the general and group factors defined by the test.* (*Coefficient of stability and equivalence*).

* This may be modified by requiring constancy over some specified period (one year, one day, etc.)

*Definition* (3): *Reliability is the degree to which the test score indicates the status of the individual at the present instant in the general and group factors defined by the test.* (*Coefficient of equivalence*). Internal consistency tests are generally measures of equivalence. These coefficients predict the correlation of the test with a hypothetical equivalent test, as like the first test as the parts of the first test are like each other.

*Definition* (4): *Reliability is the degree to which the test score indicates individual differences in any traits at the present moment.* (*Hypothetical self-correlation*).

These names are open to criticism, and better suggestions are in order. The important thing is to recognize that in the past all four of these and many approximations to them have been called "the reliability coefficient." No one of these is the "right" coefficient. They measure different things, and each is useful. What is important is to avoid confusing one with another, and using one as an estimate of another. It may be noted that reliability of a test can only be discussed in relation to a particular sample of persons.

The components of error variance under each definition imply that in practice some coefficients will be larger than others for a given test. If stability is not perfect, and if items contain some specificity loading, the hypothetical self-correlation will be greatest, and the coefficient of stability and equivalence will be the smallest of the four.

As Kelley states (7), the concept of reliability is meaningless unless one postulates that two measures of the same function exist. They may be successive measurements of a stable event, or simultaneous measurements of a unique event. But in regard to the nonrepeating event which can be observed only once, reliability has only a theoretical interest. In fact, if one accepts a deterministic position, there is no "error" in a measurement of a unique event. The student's responses and his score are determined by many forces, and we do not know what they are; but the resultant of these forces is a particular act, and the act itself, at this instant and with these particular forces, is perfectly reliable. "Chance" and "error" are merely names we give to our ignorance of what determines an event.

All methods of studying reliability make a somewhat fallacious division of variables into "real variables" and "error." It is probbaly more correct to conceive a continuum between the instantaneous behavior which has an infinitesimal period, through states of longer duration, to the virtually constant individual differences. A test score

is made up of all these "real" elements, each of which could be perfectly predicted if our knowledge were adequate. Reliability, according to this conception, becomes a measure of our ignorance of the real factors underlying brief fluctuations of behavior and atypical acts. Perhaps a new statistical method based on the non-Aristotelian conception of a continuum of realities will some day permit us to avoid the troublesome attempt to divide the continuum into "reality" and "error."

For the present, it appears to be necessary to retain the artificial separation. In thinking about the self-correlation of a test—the consistency with which it measures whatever it measures—we may class as chance effects all variables whose period of variation is shorter than the time required to take the test. Momentary fluctuations are therefore "errors," but shifts in fatigue, set, or skill having a longer cycle are possibly worth measuring.

### *Techniques of Estimation*

Each method used in the past to study "reliability" may be associated with one of these definitions. The procedures requiring more than one trial will be discussed first.

*Retest method.* The retest method calls for giving the same test twice to the same group. The trials are supposed to be independent, but this may well not be true. Shift in relative scores is always treated in the error variance, not the true-score variance; the retest coefficient is therefore an estimate of the coefficient of stability. Failure to attain independent trials may make the estimate too high or too low.

Guttman (3, 263), in a complete reconsideration of reliability theory, defines reliability in terms of the stability of individual differences during a large number of "independent" retests. He shows that the reliability thus defined (a coefficient of stability) may be estimated by the correlation between two independent trials. His definition of independence will be discussed below.

*Equivalent tests method.* Two "equivalent" or "parallel" tests may be given, with any interval between, and their correlation determined. Experimental independence is assumed, despite the effect experience with one form may have on the second. Constancy is assumed, and all shifts in relative score are treated in the error variance. Specific-factor variances are treated in the error variance. This is therefore an estimate of the coefficient of stability and equivalence. Because the assumption of independence cannot be tested, it is never

known whether the estimate is high or low. To interpret a coefficient involving equivalence, one must know how the tests are equivalent. If the tests are alike only in the general factor, group-factor variances are included as error, and the coefficient reflects the extent to which scores are determined by a stable general factor. Parallel tests should ordinarily have the same general and group factors. Were items in the two forms matched to test the same specific items of information or skill, the equivalent tests might to some degree include the same specific factors. The specific factors in the two tests could not be completely the same, however, unless the items were identical. The coefficient of equivalence is a property of a *pair* of tests and will vary according to the kind of similarity established in equating the tests. To the degree that parallel tests have the same general and group factors, the coefficient indicates the stability of performance in the general and group factors.

*The split-half method.* The widely used split-half method requires the correlation of half the items in the test with the remaining items. Cronbach has studied the effect of various splits upon the resulting coefficient (1) and has suggested the use of parallel splits, in which the two halves are made nearly equivalent (2). In the parallel split, each part represents the general factor and the group factors of the original test as well as possible. The half-tests should have equal standard deviations. The procedure makes no assumption of constancy, but does include the specific-factor variance as error variance. The split-half estimate is a coefficient of equivalence, estimating the correlation of simultaneously administered parallel tests, as like each other as are the halves of the test given. Any failure in splitting to obtain equivalent halves will tend to lower the correlation obtained. An assumption of experimental independence is made in considering the split-half correlation an estimate of the parallel-test correlation. In testing by parallel tests, the performance on one form is presumably independent of performance on the other. When items are presented together, however, there is always the possibility of spurious inter-item correlation due to item linkages and brief fluctuations of mood and attention.

Most random or odd-even splits do not represent all factors equally in both halves. If the assumption of experimental independence were valid, the correlation would therefore be an underestimate of the coefficient of equivalence. Guttman (3, 260) states that the corrected split-half coefficient is always a lower bound to "the reliability coefficient," no matter how the test is split. He cautions that this inequality is true only for an indefinitely large sample of persons. Sampling errors in practice preclude taking as one's coefficient

the largest of many trial split coefficients. Guttman defines reliability in terms of repeated independent trials of the same (not equivalent) tests. By this definition, the split-half estimate, including specificity as an error of measurement, is a low one. The coefficient of equivalence is a conservative estimate of the hypothetical self-correlation.

The assumptions of the Spearman-Brown formula have been stated in various ways, and this has led to some confusion as to the applicability of the formula. The derivation hypothecates equivalent tests and predicts their correlation from the correlation of equivalent half-tests. Equivalence is the only assumption made, and in the derivation equivalence is defined by requiring equal standard deviations of the half-tests and by requiring that the hypothetical equivalent tests be just as similar as the half-tests ($r_{ab} = r_{aA} = r_{bB} = r_{AB}$). This defines equivalence so that all tests have the same common factor composition. It makes no direct assumption of the equivalence of pairs of items or of the unit-rank among the item intercorrelations.

The items of a test may be considered as a sample of some larger population. One may define the purpose of the test in terms of the population of items to be measured; the test fulfils this purpose insofar as the items are a *representative* sample of the population. Alternatively, one may consider the test as defined by its items, and think of the population as the entire group of items of which the sample is representative. The coefficient of equivalence (obtained by the parallel-test or internal consistency methods) correlates two samples of items and indicates the extent to which the variance in each may be attributed to common factors. The extent of common-factor loadings is the extent to which test scores are determined by "the population variable." If the samples to be compared must be representative, rather than random, it is necessary, in split-half procedures, to use the parallel split or a split according to a table of specifications.

*The Kuder-Richardson formulas.* A radical reformulation of the reliability problem was offered in 1937 by Kuder and Richardson (8). They proposed several alternative formulas which have been widely adopted. The original derivation has been criticized because of the numerous assumptions made, but other writers have developed the same formulas more directly. Perhaps the simplest derivation was published by Jackson and Ferguson (5, 74). They define reliability as a coefficient of equivalence, equivalence being defined by requiring that the two tests have equal variances and that the mean inter-item covariance within each test be equal and equal to the mean inter-item covariance between tests. If these assumptions are satisfied, the Kuder-Richardson formula (20) is an exact estimate of the coefficient

of equivalence. This condition is a reasonable one when the items of a test are considered as drawn from a population of items all measuring a single general factor. If group factors are present, even though the two tests measures these group factors equally, then, $\overline{r_{ij}\, S_i S_j} < \overline{r_{ij'}\, S_i S_{j'}}$,* and the Kuder-Richardson formula gives a conservative estimate of the coefficient of equivalence—how conservative one does not know.

*The Guttman lower bounds.* The latest statement of the problem is that published by Guttman in 1945 (3). He derives six formulas for estimating a coefficient from data obtained on a single testing, all the estimates being lower than the "true reliability" if the sample is sufficiently great. His estimate $L_3$ is identical to that from Kuder-Richardson formula (20), although the derivations are dissimilar. His $L_4$ is equivalent to the split-half coefficient. $L_2$, which uses item covariances, is an original formula more difficult to compute than $L_3$ and $L_4$. $L_1$, $L_5$, and $L_6$ are expected to have little practical importance.

Guttman defines error as the variation of the score of a person over a universe of independent trials with the same test. His crucial assumption, $C_1$ (3, 265-266), defines independence so that the score of a person on any item on any trial is experimentally independent of his scores on any other items. In practice, changes in motivation, function shift, and other variables cause items administered together to vary together. Guttman classes shifts in the variables measured as errors of measurement and therefore is estimating a coefficient of stability when he demonstrates that the correlation between two independent trials on a large population may be taken as equal to "the reliability coefficient" (3, 268).

In deriving lower-bounds formulas, Guttman deals with hypothetical independent retests in which the mean covariance of two items within trials equals the mean covariance of the same items between trials. Beyond this he makes no assumption. His definition of independence requires that there be no shift in the variables measured between trials; i.e., that the hypothetical trials be simultaneous. Since he is using identical tests simultaneously, he has defined reliability as *the hypothetical self-correlation.* His formulas lead to underestimates of that coefficient.

One may study the effect on Guttman's results if his assumption of independence within trials is denied. This may occur when one item influences the answer to another by giving a clue, by causing encouragement or discouragement, or by setting up a pattern among

* i.e., the mean inter-item covariance within tests is less than the mean inter-item covariance between tests.

the responses. In the derivation of $L_1$, the assumption leads to discarding a positive covariance term from the right member of (28). As a consequence, $\lambda_1$ and $L_1$ are greater than they would be without the assumption, and may overestimate the hypothetical self-correlation as defined. In the derivation of $L_2$, $L_3$, and $L_4$, the assumption is felt in (25), where a positive covariance term is dropped from the right member. Without the assumption,

$$\gamma_{x_g x_j} > \gamma_{x_g x_j}, \quad g \neq j,$$

and the inequality given in (37) may not hold. The remainder of the derivation therefore may lead to estimates higher than the hypothetical self-correlation, if the assumption of experimental independence of items does not hold.

This weakness is common to all estimates of reliability based on a single trial. Lindquist (9, 219) points out that in the split-half method the two halves are falsely assumed to be experimentally independent, and therefore he considers the split-half estimate spuriously high. [He, however, defines reliability as what we have called the coefficient of stability and equivalence (9, 216)]. In the Kuder-Richardson formula, as derived by Jackson and Ferguson, the same assumption of independence is made when the mean inter-item covariance between tests is taken as equal to the mean covariance within tests. If motivation, response sets, and other factors common to performance on the various items of a trial are considered part of the general or group factors measured by the test, their contribution to the inter-item correlation within a trial is rightly included in the estimate of accuracy of measurement. But momentary variations which cause random changes in item covariance should not be permitted to raise the estimate obtained. Any estimate of self-correlation or equivalence based on a single trial may be higher than the hypothetical self-correlation. It may be treated as a conservative or exact estimate only if we are willing to assume that the response to each item is an independent behavior, related to response on other items only because of significant conditions in the person tested.

Guttman makes the point that his split-half formula

$$L_4 = 2\left(1 - \frac{s_1^2 + s_2^2}{s_t^2}\right) \tag{5}$$

is superior to the Spearman-Brown formula in that it does not assume the two half-tests to have equal variance. His formula can be derived as an estimate of the coefficient of equivalence, according to the usual proof of the Spearman-Brown formula, except that equivalence is de-

fined so that $\sigma_{a+b} = \sigma_{A+B}$, and $r_{aA}\sigma_a\sigma_A = r_{aB}\sigma_a\sigma_B = r_{Ab}\sigma_A\sigma_b = r_{bB}\sigma_b\sigma_B = r_{ab}\sigma_a\sigma_b$. This leads to a formula identical to Guttman's, or an equivalent form previously derived by Flanagan (see Kelley, 7) which is less readily computed. Values obtained using this formula are smaller (usually by a small amount) than the values from the Spearman-Brown formula, except where $s_a = s_b$. It appears that this formula should replace the Spearman-Brown procedure.

*Summary*

Four possible definitions of "reliability" have been considered. The hypothetical self-correlation requires independent simultaneous identical tests. For psychological variables this is a hypothetical situation, and no one has found an unbiased estimate of this coefficient. Guttman's formula $L_2$ would be a conservative estimate of the hypothetical self-correlation, save for the necessity of assuming that responses to one item are not influenced by responses to another item. Guttman's $L_2$ is ordinarily greater than the estimate from the Kuder-Richardson formula.

The coefficient of equivalence is lower than the hypothetical self-correlation. Kuder-Richardson formula (20) is an exact estimate of the coefficient of equivalence for tests where the item intercorrelation matrix has rank one; otherwise the estimate is conservative. This, however, like all estimates of equivalence, assumes experimental independence of items within one trial. The parallel-split method gives an estimate of the coefficient of equivalence. For an ideally large population, the highest split-coefficient is the best estimate, and estimates from other splits are conservative, save for the failure of independence of items.

The coefficient of stability is lower than the hypothetical self-correlation. It is estimated by the test-retest correlation, but carry-over from one test to another may cause the estimate to be faulty.

The parallel-tests correlation is an estimate of the coefficient of stability and equivalence. It may be unduly high if the two tests are not experimentally independent. Otherwise, the estimate will ordinarily be lower than the coefficient of stability or the coefficient of equivalence.

A simple table may indicate the different meanings of the various procedures. In Table 1, checks indicate the variances which are included in the error of measurement, according to each procedure. In the absence of sampling error, any estimate of reliability is less than the hypothetical self-correlation, assuming experimental independence. Every procedure assumes either the experimental independence of trials or of items within the trials. This condition is rarely

satisfied, and any obtained coefficient may therefore be higher than the coefficient supposed to be obtained.

TABLE 1
Variances Included in Error Variance of a Test, According to Various Formulations of the Reliability Problem*

| | General Factor Variance | Group Factor Variance | Specifc Factor Variance | Instability Variance, General Factor | Instability Variances, Group Factors | Instability Variances, Specific Factors | Errors of Measurement of Items |
|---|---|---|---|---|---|---|---|
| Test-Retest | | | | x | x | x | x |
| Parallel Test | | | x | x | x | x | x |
| Parallel Split | | | x | | | | x |
| Random Split | | x | x | | | | x |
| Kuder-Richardson (20) | | x | x | | | | x |
| Guttman $L_2$ | | | | | | | x† |
| Hypothetical Self-Correlation | | | | | | | x |
| Coefficient of Equivalence | | | x | | | | x |
| Coefficient of Stability | | | | x | x | x | x |
| Coefficient of Stability and Equivalence | | | x | x | x | x | x |

* An *x* indicates that the variance indicated is included in the error of measurement by the procedure or definition listed at the left.
† In equations (31) and (43), Guttman sets up inequalities which overestimate the item error variance.

### *Practical Implications*

No one "best" estimate of reliability exists. If one could validly make the assumption of stability between trials, and independence of trials, the test-retest correlation would be satisfactory. Frequently we must rely on single-trial estimates. Guttman's $L_2$ or a parallel-split used with his $L_3$ will in general give the highest coefficients. Where the test measures a single factor, the Kuder-Richardson formula (Guttman's $L_4$) should be as useful as the other two procedures.

In many situations, it is appropriate to seek a coefficient other than the hypothetical self-correlation. In correcting for attenuation, any of the coefficients described in this paper may be appropriate. Following the lead of Remmers and Whisler (11), one may distinguish between the "true instantaneous score" in a variable (related to the self-correlation or the coefficient of equivalence) and the "true score" in a trait (related to the coefficient of stability or of stability and equivalence). Sometimes one wishes to know the correlation between true scores in two traits postulated as stable over a period of time—"somatotype" vs. "temperament" is a typical problem. Here the appropriate coefficients for use in the attenuation formula are the

coefficient of stability (if the trait is defined operationally by a specific test) or the coefficient of stability and equivalence (if the trait is defined by a family of similar tests). Other problems call for studying the relation between true instantaneous score in one variable (such as an aptitude test) and true score in another defined as stable (such as job performance). For this, the reliability of the former score would be based on a coefficient of equivalence (since the hypothetical self-correlation is not known), and the reliability of the latter would be based on one of the coefficients involving stability. The third possibility, and one of much theoretical importance, is a problem regarding true instantaneous scores in two variables, such as mood and performance. The correction for attenuation here requires use of two coefficients of equivalence.

Similar reasoning applies to the problem of estimating the significance of changes in test score. If the identical test is given both times, the coefficient of stability is appropriate. The hypothetical self-correlation, if known, would test whether a significant change in behavior had occurred, although this change might be due to normal diurnal fluctuation. The coefficient of stability tests whether the change is greater than that "normally" to be expected due to function fluctuation. If growth is measured by equivalent tests, a coefficient of equivalence, or of stability and equivalence, is relevant.

*In evaluating a test, all four coefficients are of interest.* For most purposes, one wishes to measure stable characteristics, so that a coefficient of stability is needed. For research purposes, however, a test having high instantaneous self-correlation or equivalence and low stability may be very satisfactory.

The coefficient of stability is an abstraction; in reality, there is an indefinitely large number of such coefficients, corresponding to various time intervals between tests. For meaningful use of such a coefficient, it must be defined as "the coefficient of stability over one week," or the like. The coefficient also depends on the conditions affecting the subject between testings. Strictly speaking, a coefficient of stability may be carried over to a new situation only when the time interval and the conditions between testings are similar to those under which the coefficient was obtained. The coefficient of stability would be better understood if research were available showing how the coefficient varies with increasing time lapse.

The following recommendations result from the analysis made above.

1. Reliability for psychological measurement can never be observed as in the physical sciences, where variables are practically constant and non-hysteretic. All estimates of reliability require as-

sumptions unlikely to be fulfilled.

2. Several coefficients numerically less than the hypothetical self-correlation can be estimated. A distinction between these various coefficients should be made; the writer proposes the names coefficient of equivalence, coefficient of stability, and coefficient of stability and equivalence.

3. The coefficient of equivalence may be estimated by the parallel-split method, using formula (5), Guttman's $L_4$. The Kuder-Richardson formula (20) underestimates this coefficient unless the test item matrix has rank one. Guttman's $L_2$ gives an underestimate of the hypothetical self-correlation which may or may not be higher than the coefficient of equivalence. All estimates of reliability or equivalence based on a single trial assume that test items are experimentally independent. To the extent that this is untrue, estimates may be erroneously high.

4. The coefficient of stability may be estimated by the test-retest method, with an undetermined error due to failure of independence. The coefficient of stability and equivalence may be estimated by the correlation of parallel tests, with a similar error.

5. In describing a test, the author should provide separate estimates of the coefficient of equivalence and the coefficient of stability. The time interval used in obtaining the coefficient of stability should be reported. If there are multiple forms, the coefficient of stability for each should be given.

6. In practice, the coefficient of equivalence or the coefficient of stability may be used meaningfully where the reliability coefficient is called for. The coefficients are not interchangeable and have different meanings in corrections for attenuation, standard errors of measurement, and like applications. The hypothetical self-correlation, showing the extent to which a test measures real but possibly momentary differences in performance, is more important to the theory of measurement than to the practical use of tests.

## REFERENCES

1. Cronbach, L. J. A case study of the split-half reliability coefficient. *J. educ. Psychol.*, in press.
2. Cronbach, L. J. On estimates of test reliability. *J. educ. Psychol.*, 1943, **34**, 485-494.
3. Guttman, L. A basis for analyzing test-retest reliability. *Psychometrika*, 1945, **10**, 255-282.
4. Holzinger, K. J. and Harman, H. Factorial analysis. Chicago: University of Chicago Press, 1941.
5. Jackson, R. W. B. and Ferguson, G. A. Studies on the reliability of tests. Toronto: Department of Educational Research, Bulletin No. 12, 1941.
6. Jenkins, J. G. Validity for what? *J. consulting psychol.*, 1946, **10**, 93-98.

7. Kelley, T. L. The reliability coefficient. *Psychometrika*, 1942, 7, 75-83.
8. Kuder, G. F. and Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, 2, 151-160.
9. Lindquist, E. F. A first course in statistics. Boston: Houghton-Mifflin, 1942.
10. London, I. D. Some consequences for history and psychology of Langmuir's concept of convergence and divergence of phenomena. *Psychol. Rev.* 1946, 53, 170-188.
11. Remmers, H. H. and Whisler, L. Test reliability as a function of method of computation. *J. educ. Psychol.*, 1938, 29, 81-92.

PSYCHOMETRIKA—VOL. 12, NO. 1
MARCH, 1947

# TABLE FOR DETERMINING PHI COEFFICIENTS

C. E. JURGENSEN
MINNEAPOLIS GAS LIGHT COMPANY

A table is presented which directly gives phi coefficients accurate to three places when entered by the proportion of one sub-group responding in a specified manner and the proportion of a second sub-group responding in the same manner. The table gives coefficients identical with those obtained by formula if the sub-groups are equal in number. The phi coefficients can readily be expressed, if desired, in terms of critical ratio or chi square. The table is more accurate than the use of abacs and eliminates the use of time-consuming formulas. Accurate determination of item validity on the basis of statistically rigorous techniques can be made more quickly by means of the table than validity determined by less efficient methods which have previously been used to save time.

Increasing emphasis on item analyses is being found in test construction and development, particularly with regard to item validity as estimated by means of internal consistency or an outside criterion. Long and Sandiford (5) have summarized the methods which were most popular a decade ago, and test construction since that time seems to have continued to follow those methods in the main.

Methods for determining whether individual items should be retained or rejected (or the closely allied problem of determining what weights should be assigned each item) vary from a simple determination of the difference in per cent of persons in contrasted groups who respond similarly to a given item, to the more rigorous correlational techniques or levels of significance. Because of the great amount of computational time necessary to determine item validities on the basis of correlational techniques, the simpler and less efficient techniques have too often been used. This is particularly unfortunate in those cases where the number of cases is small and has often resulted in retaining items purely on the basis of chance differences. In an effort to reduce the time required for the more accurate types of item validation, various tables, nomographs, and abacs have been developed.

One of the earliest of these techniques was a table developed by Edgerton and Paterson (1) which gives standard errors from .001 to.100 by successive steps of .001 for all possible combinations of differences between percentages. For a maximum percentage difference of 50, this table covers a maximum standard error for 25 cases

and a minimum standard error for one million cases. These data do not give item validity, but do permit necessary computations to be made more readily than otherwise.

Votaw (8) has published the formulas necessary to construct an abac based on the probable error of differences between two groups as well as an abac based on a critical ratio of 2 (in terms of probable error) when the $N$ of each subgroup equals 45. Mosier and McQuitty (7) have published a more detailed abac giving critical ratios of 2, 3, 5, 7 and 10 (in terms of standard error) based on the highest and lowest fifty cases of a group having a total $N$ of 200. Mosier and McQuitty also published correlational abacs based on upper-lower halves and upper-lower quarters. Guilford (4) has published an abac based on phi coefficients ranging from 0 to $\pm.90$ in steps of .10 and another giving 1% and 5% levels of significance when the total $N$ equals 50, 100, 200, and 400.

Lord (6) has given an alignment chart for calculating the fourfold point correlation coefficient, the chart being entered on three values: per cent of cases successful with respect to the first variable, the similar per cent for the second variable, and the per cent of cases successful with respect to both variables.

Fiske and Dunlap (2) have published a formula for constructing an ellipse on the assumption that the two sub-groups are random samples from the same parent population and that the best estimate of the true proportion is the weighted mean proportion of the two samples. Fiske and Dunlap's abac is based on a critical ratio of 2 with one hundred persons in each sub-group.

Numerous other abacs and nomographs have been constructed. Although they all possess the advantage of reducing statistical computations, other objections are inherent in their use. Abacs and nomographs based on critical ratios, level of significance, or chi square are necessarily constructed for use in situations having a specified number of cases in each sub-group. As the $N$ is changed, so must the abac be changed also. Inasmuch as construction of an abac requires computing numerous points by means of formula and careful drawing of a curved line to fit these points, and because the number of abacs which can be drawn is unlimited, this procedure becomes impractical.

Rather than constructing an abac for each possible $N$, it is also possible to devise abacs for various selected $N$'s, and in any single study the abac can be used which most closely approximates the available data. Although such approximations are sufficiently accurate for most practical purposes, many research workers are reluctant to state in the literature that their work is based on approximations

lest such statement be interpreted as an indication of flatulent work.

Another possible procedure is to discard data to the extent that the $N$ fits one of the available abacs. This procedure may be satisfactory when data are based on several hundred cases but cannot be recommended when the original $N$ is small. Necessity frequently dictates a small $N$, and it is inadvisable to further reduce the $N$ in order to permit quicker analysis of data by means of an available abac.

Another difficulty inherent in abacs is the difficulty of accurately determining exact item validity when interpolations must be made between the ellipses of an abac. The error of such estimates may be further increased by the necessity for interpolation at the points of entry on the abac.

Some of the objections to nomographs and abacs can be overcome by using a table which is entered in a similar manner; namely, the proportion of one sub-group responding in a specified way supplies the vertical entry and the proportion of another sub-group responding in the same way is entered in the horizontal dimension. Exact item validity can then be read at the point of intersection of the column and row.

In order to be of maximum value, a table of this type should be expressed in terms which permit its use with any desired number of cases and should be such as to permit rapid determination of the degree of significance of any difference. The table of phi coefficients accompanying this article fulfills these conditions. The following interrelationships obtain when the number of cases in the two sub-groups is equal:

$$\text{Phi Coefficient } (\phi) = \sqrt{\frac{\chi^2}{N_{tot}}} = \frac{CR}{\sqrt{N_{tot}}}; \tag{1}$$

$$\text{Critical Ratio } (CR) = \phi\sqrt{N_{tot}} = \sqrt{\chi^2}; \tag{2}$$

$$\text{Chi Square } (\chi^2) = N\phi^2 = CR^2. \tag{3}$$

A Pearson $r$ corresponding to the phi coefficient can be estimated if desired. If both variables can be considered as being continuous, the Pearson $r$ is estimated by dividing $\phi$ by .637. If one set of data are genuinely dichotomous and the other is continuous but artificially reduced to a dichotomy, the Pearson $r$ can be estimated by dividing $\phi$ by .798, although, as Guilford (3) points out, the meaning of such figures is questionable and interpretation should be made only with extreme caution and with cognizance of the steps by which the coefficient was derived. If true point distributions are involved $\phi$ is numerically equivalent to the Pearson $r$.

In addition to being readily converted to $CR$ and $\chi^2$, the phi coefficient has the added advantages of being widely applicable in many situations, and being one of the few coefficients which can properly be used in some of these situations.

Assuming that the sub-groups are equal in size, the formula for $\phi$ is usually expressed as:

$$\phi = \frac{p_u - p_l}{\sqrt{pq}}, \tag{4}$$

where $p_u$ = the proportion in terms of total $N$ of one sub-group (upper) responding in a specified way,

$p_l$ = the proportion of the other sub-group (lower) responding in the same way,

$p$ = the total proportion $(p_u + p_l)$ responding in the specified way,

$q$ = the total not responding in the specified way $(1.00 - p)$.

In item analysis the test constructor usually deals with proportions expressed in terms of each sub-group rather than the total $N$. In such case $p + q = 2.00$. The formula can then be expressed as:

$$\phi = \frac{\frac{p_u}{2} - \frac{p_l}{2}}{\sqrt{\frac{p}{2} \cdot \frac{q}{2}}}, \tag{5}$$

where $p = p_u + p_l$ and $q = 2 - (p_u - p_l)$.

Formula (5) can be expressed entirely in terms of $p_u$ and $p_l$ as:

$$\phi = \frac{p_u - p_l}{\sqrt{(p_u + p_l)(2 - p_u - p_l)}}. \tag{6}$$

Formula (6) was used in constructing the accompanying table. The 200 coefficients on the outer edges of the table were computed by formula to be accurate to six decimal places. The 4850 remaining coefficients were obtained by successive subtraction of a constant from each of the outer-edge entries. The constant results from the fact that $p_u + p_l$ remains the same and $p_u - p_l$ decreases systematically by .02 on any diagonal from the outer edge which is perpendicular to the center diagonal which separates the table into positive and negative coefficients. For example: Select a $p_u$ of .71 and $p_l$ of .00, which appears on the outer edge of the table. On a diagonal perpendicular to the center dividing line, $p_u$ and $p_l$ progressively become .70 and .01, .69 and .02, .68 and .03, .67 and .04, etc. In each case $p_u + p_l = .71$, and $p_u - p_l$ progressively decreases by .02 from .71 to .69, .67,

.65, .63, etc. The subtracted constant was therefore obtained by multiplying the reciprocal of the denominator of formula (6) by .02. Successive subtractions from the outer-edge phi coefficient computed by formula thus gave each of the other entries in the diagonal.

Several checks were made to insure accuracy of the table: (1) Each of the two hundred outer-edge phi coefficients was separately computed by two persons, (2) each of the two hundred subtractive constants was computed separately by two persons, (3) original phi coefficients and subtractive constants were checked on the basis of comparable quadrants of the table, and (4) each row and each column of the final table were checked separately by two persons on the basis of comparable quadrants.

The table of phi coefficients is simple to use. The proportion of each of two sub-groups responding to an item in a specified manner is determined, the proportions being computed on the basis of the number of cases within each sub-group rather than the number of cases within the total group. The table is entered vertically by one of the proportions and horizontally by the other. The phi coefficient is given at the point of intersection of the row and column entries. For purposes of reducing the size of the table, only the positive coefficients are included. If entry in the usual manner leads to a blank cell in the table, the coefficient can be found by reversing the horizontal and the vertical entry. The manner of entry together with the type of data being handled readily indicates whether the coefficient is positive or negative.

The table assumes an equal number of cases in the two sub-groups, and if this assumption is met the coefficient is the same as a coefficient computed by formula. The greater the difference between the *N*'s of the two sub-groups the greater will the table coefficient differ from the computed coefficient.

Knowledge on the part of the user of the number of cases included in the groups will permit a quick determination of critical ratio, level of significance, or chi square as given by formulas (1), (2), and (3). The user can then set his own standards for accepting or rejecting items, and for assigning weights to items.

Inasmuch as this table permits rapid and accurate determination of item validity on the basis of statistically rigorous techniques, there is no justification for using less efficient methods. Such inefficient methods now require as much time as the more acceptable, but hitherto time-consuming methods.

## REFERENCES

1. Edgerton, H. A. and Paterson, D. G. Table of standard errors and probable errors of percentages for varying numbers of cases. *J. appl. Psychol.*, 1926, 10, 378-391.

2. Fiske, D. W. and Dunlap, J. W. A graphical test for the significance of differences between frequencies from different samples. *Psychometrika*, 1945, 10, 225-227.
3. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1942, p. 247.
4. Guilford, J. P. The phi coefficient and chi square as indices of item validity. *Psychometrika*, 1941, 6, 11-19.
5. Long, J. A. and Sandiford, P. *The validation of test items.* Bulletin No. 3, Department of Educational Research. Toronto: University of Toronto, 1935. Pp. 126.
6. Lord, F. M. Alignment chart for calculating the fourfold point correlation coefficient. *Psychometrika*, 1944, 9, 41-42.
7. Mosier, C. I. and McQuitty, J. V. Methods of item validation and abacs for item-test correlation and critical ratio of upper-lower differences. *Psychometrika*, 1940, 5, 57-65.
8. Votaw, D. F. Graphical determination of probable errors in validation of test items. *J. educ. Psychol.*, 1933, 24, 682-686.

| | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 1000 | 990 | 980 | 970 | 961 | 951 | 942 | 932 | 923 | 914 | 905 | 895 | 886 | 877 | 869 | 860 | 851 | 842 | 834 | 825 |
| 99 | 990 | 980 | 970 | 960 | 950 | 941 | 931 | 922 | 912 | 903 | 894 | 884 | 875 | 866 | 857 | 848 | 839 | 831 | 822 | 813 |
| 98 | 980 | 970 | 960 | 950 | 940 | 930 | 921 | 911 | 902 | 892 | 883 | 874 | 864 | 855 | 846 | 837 | 828 | 819 | 810 | 802 |
| 97 | 970 | 960 | 950 | 940 | 930 | 920 | 910 | 901 | 891 | 882 | 872 | 863 | 853 | 844 | 835 | 826 | 817 | 808 | 799 | 790 |
| 96 | 961 | 950 | 940 | 930 | 920 | 910 | 900 | 890 | 881 | 871 | 862 | 852 | 843 | 833 | 824 | 815 | 806 | 797 | 788 | 779 |
| 95 | 951 | 941 | 930 | 920 | 910 | 900 | 890 | 880 | 870 | 861 | 851 | 842 | 832 | 823 | 813 | 804 | 795 | 786 | 777 | 768 |
| 94 | 942 | 931 | 921 | 910 | 900 | 890 | 880 | 870 | 860 | 850 | 841 | 831 | 821 | 812 | 803 | 793 | 784 | 775 | 766 | 756 |
| 93 | 932 | 922 | 911 | 901 | 890 | 880 | 870 | 860 | 850 | 840 | 830 | 821 | 811 | 801 | 792 | 783 | 773 | 764 | 755 | 745 |
| 92 | 923 | 912 | 902 | 891 | 881 | 870 | 860 | 850 | 840 | 830 | 820 | 810 | 801 | 791 | 781 | 772 | 762 | 753 | 744 | 734 |
| 91 | 914 | 903 | 892 | 882 | 871 | 861 | 850 | 840 | 830 | 820 | 810 | 800 | 790 | 781 | 771 | 761 | 752 | 742 | 733 | 724 |
| 90 | 905 | 894 | 883 | 872 | 862 | 851 | 841 | 830 | 820 | 810 | 800 | 790 | 780 | 770 | 761 | 751 | 741 | 732 | 722 | 713 |
| 89 | 895 | 884 | 874 | 863 | 852 | 842 | 831 | 821 | 810 | 800 | 790 | 780 | 770 | 760 | 750 | 741 | 731 | 721 | 712 | 702 |
| 88 | 886 | 875 | 864 | 853 | 843 | 832 | 821 | 811 | 801 | 790 | 780 | 770 | 760 | 750 | 740 | 730 | 721 | 711 | 701 | 692 |
| 87 | 877 | 866 | 855 | 844 | 833 | 823 | 812 | 801 | 791 | 781 | 770 | 760 | 750 | 740 | 730 | 720 | 710 | 701 | 691 | 681 |
| 86 | 869 | 857 | 846 | 835 | 824 | 813 | 803 | 792 | 781 | 771 | 761 | 750 | 740 | 730 | 720 | 710 | 700 | 690 | 681 | 671 |
| 85 | 860 | 848 | 837 | 826 | 815 | 804 | 793 | 783 | 772 | 761 | 751 | 741 | 730 | 720 | 710 | 700 | 690 | 680 | 670 | 661 |
| 84 | 851 | 839 | 828 | 817 | 806 | 795 | 784 | 773 | 762 | 752 | 741 | 731 | 721 | 710 | 700 | 690 | 680 | 670 | 660 | 650 |
| 83 | 842 | 831 | 819 | 808 | 797 | 786 | 775 | 764 | 753 | 742 | 732 | 721 | 711 | 701 | 690 | 680 | 670 | 660 | 650 | 640 |
| 82 | 834 | 822 | 810 | 799 | 788 | 777 | 766 | 755 | 744 | 733 | 722 | 712 | 701 | 691 | 681 | 670 | 660 | 650 | 640 | 630 |
| 81 | 825 | 813 | 802 | 790 | 779 | 768 | 756 | 745 | 734 | 724 | 713 | 702 | 692 | 681 | 671 | 661 | 650 | 640 | 630 | 620 |
| 80 | 816 | 805 | 793 | 781 | 770 | 759 | 747 | 736 | 725 | 714 | 704 | 693 | 682 | 672 | 661 | 651 | 641 | 630 | 620 | 610 |
| 79 | 808 | 796 | 784 | 773 | 761 | 750 | 738 | 727 | 716 | 705 | 694 | 683 | 673 | 662 | 652 | 641 | 631 | 620 | 610 | 600 |
| 78 | 800 | 788 | 776 | 764 | 752 | 741 | 729 | 718 | 707 | 696 | 685 | 674 | 663 | 653 | 642 | 632 | 621 | 611 | 600 | 590 |
| 77 | 791 | 779 | 767 | 755 | 744 | 732 | 720 | 709 | 698 | 687 | 676 | 665 | 654 | 643 | 633 | 622 | 612 | 601 | 591 | 580 |
| 76 | 783 | 771 | 759 | 747 | 735 | 723 | 712 | 700 | 689 | 678 | 667 | 656 | 645 | 634 | 623 | 612 | 602 | 591 | 581 | 571 |
| 75 | 775 | 762 | 750 | 738 | 726 | 714 | 703 | 691 | 680 | 669 | 657 | 646 | 635 | 625 | 614 | 603 | 592 | 582 | 571 | 561 |
| 74 | 766 | 754 | 742 | 730 | 718 | 706 | 694 | 682 | 671 | 660 | 648 | 637 | 626 | 615 | 604 | 594 | 583 | 572 | 562 | 551 |
| 73 | 758 | 746 | 733 | 721 | 709 | 697 | 685 | 674 | 662 | 651 | 639 | 628 | 617 | 606 | 595 | 584 | 573 | 563 | 552 | 542 |
| 72 | 750 | 737 | 725 | 713 | 700 | 688 | 677 | 665 | 653 | 642 | 630 | 619 | 608 | 597 | 586 | 575 | 564 | 553 | 543 | 532 |
| 71 | 742 | 729 | 717 | 704 | 692 | 680 | 668 | 656 | 644 | 633 | 621 | 610 | 599 | 588 | 577 | 566 | 555 | 544 | 533 | 523 |
| 70 | 734 | 721 | 708 | 696 | 684 | 671 | 659 | 647 | 636 | 624 | 612 | 601 | 590 | 578 | 567 | 556 | 545 | 535 | 524 | 513 |
| 69 | 726 | 713 | 700 | 688 | 675 | 663 | 651 | 639 | 627 | 615 | 603 | 592 | 581 | 569 | 558 | 547 | 536 | 525 | 514 | 504 |
| 68 | 718 | 705 | 692 | 679 | 667 | 654 | 642 | 630 | 618 | 606 | 595 | 583 | 572 | 560 | 549 | 538 | 527 | 516 | 505 | 494 |
| 67 | 710 | 697 | 684 | 671 | 658 | 646 | 634 | 621 | 609 | 597 | 586 | 574 | 563 | 551 | 540 | 529 | 518 | 507 | 496 | 485 |
| 66 | 702 | 689 | 676 | 663 | 650 | 637 | 625 | 613 | 601 | 589 | 577 | 565 | 554 | 542 | 531 | 519 | 508 | 497 | 486 | 475 |
| 65 | 694 | 681 | 667 | 654 | 642 | 629 | 616 | 604 | 592 | 580 | 568 | 556 | 545 | 533 | 522 | 510 | 499 | 488 | 477 | 466 |
| 64 | 686 | 673 | 659 | 646 | 633 | 621 | 608 | 596 | 583 | 571 | 559 | 547 | 536 | 524 | 513 | 501 | 490 | 479 | 468 | 457 |
| 63 | 678 | 665 | 651 | 638 | 625 | 612 | 600 | 587 | 575 | 563 | 550 | 539 | 527 | 515 | 503 | 492 | 481 | 469 | 458 | 447 |
| 62 | 670 | 657 | 643 | 630 | 617 | 604 | 591 | 579 | 566 | 554 | 542 | 530 | 518 | 506 | 494 | 483 | 472 | 460 | 449 | 438 |
| 61 | 662 | 649 | 635 | 622 | 608 | 595 | 583 | 570 | 557 | 545 | 533 | 521 | 509 | 497 | 485 | 474 | 462 | 451 | 440 | 429 |
| 60 | 655 | 641 | 627 | 614 | 600 | 587 | 574 | 561 | 549 | 536 | 524 | 512 | 500 | 488 | 476 | 465 | 453 | 442 | 431 | 419 |
| 59 | 647 | 633 | 619 | 605 | 592 | 579 | 566 | 553 | 540 | 528 | 515 | 503 | 491 | 479 | 467 | 456 | 444 | 433 | 421 | 410 |
| 58 | 639 | 625 | 611 | 597 | 584 | 570 | 557 | 544 | 532 | 519 | 507 | 494 | 482 | 470 | 458 | 447 | 435 | 423 | 412 | 401 |
| 57 | 631 | 617 | 603 | 589 | 576 | 562 | 549 | 536 | 523 | 510 | 498 | 486 | 473 | 461 | 449 | 438 | 426 | 414 | 403 | 391 |
| 56 | 624 | 609 | 595 | 581 | 567 | 554 | 541 | 527 | 514 | 502 | 489 | 477 | 464 | 452 | 440 | 428 | 417 | 405 | 394 | 382 |
| 55 | 616 | 601 | 587 | 573 | 559 | 546 | 532 | 519 | 506 | 493 | 480 | 468 | 456 | 443 | 431 | 419 | 408 | 396 | 384 | 373 |
| 54 | 608 | 593 | 579 | 565 | 551 | 537 | 524 | 510 | 497 | 484 | 472 | 459 | 447 | 434 | 422 | 410 | 398 | 387 | 375 | 364 |
| 53 | 600 | 586 | 571 | 557 | 543 | 529 | 515 | 502 | 489 | 476 | 463 | 450 | 438 | 425 | 413 | 401 | 389 | 377 | 366 | 354 |
| 52 | 593 | 578 | 563 | 549 | 535 | 521 | 507 | 493 | 480 | 467 | 454 | 441 | 429 | 416 | 404 | 392 | 380 | 368 | 356 | 345 |
| 51 | 585 | 570 | 555 | 541 | 526 | 512 | 498 | 485 | 471 | 458 | 445 | 432 | 420 | 407 | 395 | 383 | 371 | 359 | 347 | 335 |
| 50 | 577 | 562 | 547 | 532 | 518 | 504 | 490 | 476 | 463 | 450 | 436 | 424 | 411 | 398 | 386 | 374 | 362 | 350 | 338 | 326 |
| | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |

| | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 49 | 570 | 554 | 539 | 524 | 510 | 496 | 482 | 468 | 454 | 441 | 428 | 415 | 402 | 389 | 377 | 364 | 352 | 340 | 328 | 317 |
| 48 | 562 | 546 | 531 | 516 | 502 | 487 | 473 | 459 | 445 | 432 | 419 | 406 | 393 | 380 | 368 | 355 | 343 | 331 | 319 | 307 |
| 47 | 554 | 539 | 523 | 508 | 493 | 479 | 464 | 450 | 437 | 423 | 410 | 397 | 384 | 371 | 358 | 346 | 334 | 322 | 310 | 298 |
| 46 | 547 | 531 | 515 | 500 | 485 | 470 | 456 | 442 | 428 | 414 | 401 | 388 | 375 | 362 | 349 | 337 | 324 | 312 | 300 | 288 |
| 45 | 539 | 523 | 507 | 492 | 477 | 462 | 447 | 433 | 419 | 405 | 392 | 379 | 366 | 353 | 340 | 327 | 315 | 303 | 291 | 279 |
| 44 | 531 | 515 | 499 | 483 | 468 | 453 | 439 | 424 | 410 | 397 | 383 | 370 | 356 | 343 | 331 | 318 | 306 | 293 | 281 | 269 |
| 43 | 523 | 507 | 491 | 475 | 460 | 445 | 430 | 416 | 402 | 388 | 374 | 360 | 347 | 334 | 321 | 309 | 296 | 284 | 271 | 259 |
| 42 | 516 | 499 | 483 | 467 | 451 | 436 | 421 | 407 | 393 | 379 | 365 | 351 | 338 | 325 | 312 | 299 | 286 | 274 | 262 | 250 |
| 41 | 508 | 491 | 475 | 459 | 443 | 428 | 413 | 398 | 384 | 370 | 356 | 342 | 329 | 315 | 302 | 290 | 277 | 264 | 252 | 240 |
| 40 | 500 | 483 | 466 | 450 | 435 | 419 | 404 | 389 | 375 | 360 | 346 | 333 | 319 | 306 | 293 | 280 | 267 | 255 | 242 | 230 |
| 39 | 492 | 475 | 458 | 442 | 426 | 410 | 395 | 380 | 366 | 351 | 337 | 323 | 310 | 296 | 283 | 270 | 258 | 245 | 233 | 220 |
| 38 | 484 | 467 | 450 | 433 | 417 | 402 | 386 | 371 | 356 | 342 | 328 | 314 | 300 | 287 | 274 | 261 | 248 | 235 | 223 | 210 |
| 37 | 476 | 459 | 442 | 425 | 409 | 393 | 377 | 362 | 347 | 333 | 318 | 304 | 291 | 277 | 264 | 251 | 238 | 225 | 213 | 200 |
| 36 | 469 | 451 | 433 | 416 | 400 | 384 | 368 | 353 | 338 | 323 | 309 | 295 | 281 | 267 | 254 | 241 | 228 | 215 | 203 | 190 |
| 35 | 461 | 442 | 425 | 408 | 391 | 375 | 359 | 344 | 329 | 314 | 299 | 285 | 271 | 258 | 244 | 231 | 218 | 205 | 193 | 180 |
| 34 | 453 | 434 | 416 | 399 | 382 | 366 | 350 | 334 | 319 | 304 | 290 | 275 | 261 | 248 | 234 | 221 | 208 | 195 | 182 | 170 |
| 33 | 445 | 426 | 408 | 390 | 373 | 357 | 341 | 325 | 310 | 295 | 280 | 266 | 251 | 238 | 224 | 211 | 198 | 185 | 172 | 160 |
| 32 | 436 | 418 | 399 | 382 | 364 | 348 | 331 | 315 | 300 | 285 | 270 | 256 | 241 | 228 | 214 | 200 | 187 | 174 | 162 | 149 |
| 31 | 428 | 409 | 391 | 373 | 355 | 338 | 322 | 306 | 290 | 275 | 260 | 246 | 231 | 217 | 204 | 190 | 177 | 164 | 151 | 139 |
| 30 | 420 | 401 | 382 | 364 | 346 | 329 | 312 | 296 | 280 | 265 | 250 | 235 | 221 | 207 | 193 | 180 | 166 | 153 | 140 | 128 |
| 29 | 412 | 392 | 373 | 355 | 337 | 319 | 303 | 286 | 270 | 255 | 240 | 225 | 211 | 196 | 183 | 169 | 156 | 143 | 130 | 117 |
| 28 | 403 | 383 | 364 | 345 | 327 | 310 | 293 | 276 | 260 | 245 | 229 | 215 | 200 | 186 | 172 | 158 | 145 | 132 | 119 | 106 |
| 27 | 395 | 375 | 355 | 336 | 318 | 300 | 283 | 266 | 250 | 234 | 219 | 204 | 189 | 175 | 161 | 147 | 134 | 121 | 108 | 095 |
| 26 | 387 | 366 | 346 | 327 | 308 | 290 | 273 | 256 | 240 | 224 | 208 | 193 | 178 | 164 | 150 | 136 | 123 | 110 | 097 | 084 |
| 25 | 378 | 357 | 337 | 317 | 298 | 280 | 262 | 245 | 229 | 213 | 197 | 182 | 167 | 153 | 139 | 125 | 111 | 098 | 085 | 072 |
| 24 | 369 | 348 | 327 | 307 | 288 | 270 | 252 | 235 | 218 | 202 | 186 | 171 | 156 | 142 | 127 | 114 | 100 | 087 | 074 | 061 |
| 23 | 360 | 339 | 317 | 297 | 278 | 259 | 241 | 224 | 207 | 191 | 175 | 160 | 145 | 130 | 116 | 102 | 088 | 075 | 062 | 049 |
| 22 | 352 | 329 | 308 | 287 | 268 | 249 | 231 | 213 | 196 | 180 | 164 | 148 | 133 | 118 | 104 | 090 | 076 | 063 | 050 | 037 |
| 21 | 343 | 320 | 298 | 277 | 257 | 238 | 219 | 202 | 185 | 168 | 152 | 136 | 121 | 106 | 092 | 078 | 064 | 051 | 038 | 025 |
| 20 | 333 | 310 | 288 | 266 | 246 | 227 | 208 | 190 | 173 | 156 | 140 | 124 | 109 | 094 | 080 | 066 | 052 | 039 | 025 | 013 |
| 19 | 324 | 300 | 277 | 256 | 235 | 215 | 197 | 178 | 161 | 144 | 128 | 112 | 097 | 082 | 067 | 053 | 039 | 026 | 013 | 0 |
| 18 | 314 | 290 | 267 | 245 | 224 | 204 | 185 | 166 | 149 | 132 | 115 | 099 | 084 | 069 | 055 | 040 | 027 | 013 | 0 | |
| 17 | 305 | 280 | 256 | 233 | 212 | 192 | 172 | 154 | 136 | 119 | 102 | 086 | 071 | 056 | 041 | 027 | 013 | 0 | | |
| 16 | 295 | 269 | 245 | 222 | 200 | 179 | 160 | 141 | 123 | 106 | 089 | 073 | 058 | 043 | 028 | 014 | 0 | | | |
| 15 | 285 | 258 | 233 | 210 | 188 | 167 | 147 | 128 | 110 | 092 | 076 | 059 | 044 | 029 | 014 | 0 | | | | |
| 14 | 274 | 247 | 221 | 197 | 175 | 153 | 133 | 114 | 096 | 078 | 062 | 045 | 030 | 015 | 0 | | | | | |
| 13 | 264 | 235 | 209 | 184 | 161 | 140 | 119 | 100 | 082 | 064 | 047 | 031 | 015 | 0 | | | | | | |
| 12 | 253 | 223 | 196 | 171 | 147 | 125 | 105 | 085 | 067 | 049 | 032 | 016 | 0 | | | | | | | |
| 11 | 241 | 211 | 183 | 157 | 133 | 111 | 090 | 070 | 051 | 033 | 016 | 0 | | | | | | | | |
| 10 | 229 | 197 | 168 | 142 | 118 | 095 | 074 | 054 | 035 | 017 | 0 | | | | | | | | | |
| 09 | 217 | 184 | 154 | 126 | 101 | 078 | 057 | 037 | 018 | 0 | | | | | | | | | | |
| 08 | 204 | 169 | 138 | 110 | 084 | 061 | 039 | 019 | 0 | | | | | | | | | | | |
| 07 | 190 | 153 | 121 | 092 | 066 | 042 | 020 | 0 | | | | | | | | | | | | |
| 06 | 176 | 136 | 102 | 072 | 046 | 022 | 0 | | | | | | | | | | | | | |
| 05 | 160 | 117 | 082 | 051 | 024 | 0 | | | | | | | | | | | | | | |
| 04 | 143 | 096 | 059 | 027 | 0 | | | | | | | | | | | | | | | |
| 03 | 123 | 071 | 032 | 0 | | | | | | | | | | | | | | | | |
| 02 | 101 | 041 | 0 | | | | | | | | | | | | | | | | | |
| 01 | 071 | 0 | | | | | | | | | | | | | | | | | | |
| 00 | 0 | | | | | | | | | | | | | | | | | | | |
| | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |

| | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 816 | 808 | 800 | 791 | 783 | 775 | 766 | 758 | 750 | 742 | 734 | 726 | 718 | 710 | 702 | 694 | 686 | 678 | 670 | 662 |
| 99 | 805 | 796 | 788 | 779 | 771 | 762 | 754 | 746 | 737 | 729 | 721 | 713 | 705 | 697 | 689 | 681 | 673 | 665 | 657 | 649 |
| 98 | 793 | 784 | 776 | 767 | 759 | 750 | 742 | 733 | 725 | 717 | 708 | 700 | 692 | 684 | 676 | 667 | 659 | 651 | 643 | 635 |
| 97 | 781 | 773 | 764 | 755 | 747 | 738 | 730 | 721 | 713 | 704 | 696 | 688 | 679 | 671 | 663 | 654 | 646 | 638 | 630 | 622 |
| 96 | 770 | 761 | 752 | 744 | 735 | 726 | 718 | 709 | 700 | 692 | 684 | 675 | 667 | 658 | 650 | 642 | 633 | 625 | 617 | 608 |
| 95 | 759 | 750 | 741 | 732 | 723 | 714 | 706 | 697 | 688 | 680 | 671 | 663 | 654 | 646 | 637 | 629 | 621 | 612 | 604 | 595 |
| 94 | 747 | 738 | 729 | 720 | 712 | 703 | 694 | 685 | 677 | 668 | 659 | 651 | 642 | 634 | 625 | 616 | 608 | 600 | 591 | 583 |
| 93 | 736 | 727 | 718 | 709 | 700 | 691 | 682 | 674 | 665 | 656 | 647 | 639 | 630 | 621 | 613 | 604 | 596 | 587 | 579 | 570 |
| 92 | 725 | 716 | 707 | 698 | 689 | 680 | 671 | 662 | 653 | 644 | 636 | 627 | 618 | 609 | 601 | 592 | 583 | 575 | 566 | 557 |
| 91 | 714 | 705 | 696 | 687 | 678 | 669 | 660 | 651 | 642 | 633 | 624 | 615 | 606 | 597 | 589 | 580 | 571 | 563 | 554 | 545 |
| 90 | 704 | 694 | 685 | 676 | 667 | 657 | 648 | 639 | 630 | 621 | 612 | 603 | 595 | 586 | 577 | 568 | 559 | 550 | 542 | 533 |
| 89 | 693 | 683 | 674 | 665 | 656 | 646 | 637 | 628 | 619 | 610 | 601 | 592 | 583 | 574 | 565 | 556 | 547 | 539 | 530 | 521 |
| 88 | 682 | 673 | 663 | 654 | 645 | 635 | 626 | 617 | 608 | 599 | 590 | 581 | 572 | 563 | 554 | 545 | 536 | 527 | 518 | 509 |
| 87 | 672 | 662 | 653 | 643 | 634 | 625 | 615 | 606 | 597 | 588 | 578 | 569 | 560 | 551 | 542 | 533 | 524 | 515 | 506 | 497 |
| 86 | 661 | 652 | 642 | 633 | 623 | 614 | 604 | 595 | 586 | 577 | 567 | 558 | 549 | 540 | 531 | 522 | 513 | 503 | 494 | 485 |
| 85 | 651 | 641 | 632 | 622 | 612 | 603 | 594 | 584 | 575 | 566 | 556 | 547 | 538 | 529 | 519 | 510 | 501 | 492 | 483 | 474 |
| 84 | 641 | 631 | 621 | 612 | 602 | 592 | 583 | 573 | 564 | 555 | 545 | 536 | 527 | 518 | 508 | 499 | 490 | 481 | 472 | 462 |
| 83 | 630 | 620 | 611 | 601 | 591 | 582 | 572 | 563 | 553 | 544 | 535 | 525 | 516 | 507 | 497 | 488 | 479 | 469 | 460 | 451 |
| 82 | 620 | 610 | 600 | 591 | 581 | 571 | 562 | 552 | 543 | 533 | 524 | 514 | 505 | 496 | 486 | 477 | 468 | 458 | 449 | 440 |
| 81 | 610 | 600 | 590 | 580 | 571 | 561 | 551 | 542 | 532 | 523 | 513 | 504 | 494 | 485 | 475 | 466 | 457 | 447 | 438 | 429 |
| 80 | 600 | 590 | 580 | 570 | 560 | 551 | 541 | 531 | 522 | 512 | 503 | 493 | 483 | 474 | 465 | 455 | 446 | 436 | 427 | 418 |
| 79 | 590 | 580 | 570 | 560 | 550 | 540 | 531 | 521 | 511 | 502 | 492 | 482 | 473 | 463 | 454 | 444 | 435 | 425 | 416 | 407 |
| 78 | 580 | 570 | 560 | 550 | 540 | 530 | 520 | 511 | 501 | 491 | 482 | 472 | 462 | 453 | 443 | 434 | 424 | 415 | 405 | 396 |
| 77 | 570 | 560 | 550 | 540 | 530 | 520 | 510 | 500 | 491 | 481 | 471 | 461 | 452 | 442 | 433 | 423 | 414 | 404 | 394 | 385 |
| 76 | 560 | 550 | 540 | 530 | 520 | 510 | 500 | 490 | 480 | 471 | 461 | 451 | 441 | 432 | 422 | 413 | 403 | 393 | 384 | 374 |
| 75 | 551 | 540 | 530 | 520 | 510 | 500 | 490 | 480 | 470 | 460 | 451 | 441 | 431 | 421 | 412 | 402 | 392 | 383 | 373 | 364 |
| 74 | 541 | 531 | 520 | 510 | 500 | 490 | 480 | 470 | 460 | 450 | 440 | 431 | 421 | 411 | 401 | 392 | 382 | 372 | 363 | 353 |
| 73 | 531 | 521 | 511 | 500 | 490 | 480 | 470 | 460 | 450 | 440 | 430 | 420 | 411 | 401 | 391 | 381 | 371 | 362 | 352 | 342 |
| 72 | 522 | 511 | 501 | 491 | 480 | 470 | 460 | 450 | 440 | 430 | 420 | 410 | 400 | 390 | 381 | 371 | 361 | 351 | 342 | 332 |
| 71 | 512 | 502 | 491 | 481 | 471 | 460 | 450 | 440 | 430 | 420 | 410 | 400 | 390 | 380 | 370 | 361 | 351 | 341 | 331 | 322 |
| 70 | 503 | 492 | 482 | 471 | 461 | 451 | 440 | 430 | 420 | 410 | 400 | 390 | 380 | 370 | 360 | 350 | 341 | 331 | 321 | 311 |
| 69 | 493 | 482 | 472 | 461 | 451 | 441 | 431 | 420 | 410 | 400 | 390 | 380 | 370 | 360 | 350 | 340 | 330 | 321 | 311 | 301 |
| 68 | 483 | 473 | 462 | 452 | 441 | 431 | 421 | 411 | 400 | 390 | 380 | 370 | 360 | 350 | 340 | 330 | 320 | 310 | 301 | 291 |
| 67 | 474 | 463 | 453 | 442 | 432 | 421 | 411 | 401 | 390 | 380 | 370 | 360 | 350 | 340 | 330 | 320 | 310 | 300 | 290 | 281 |
| 66 | 465 | 454 | 443 | 433 | 422 | 412 | 401 | 391 | 381 | 370 | 360 | 350 | 340 | 330 | 320 | 310 | 300 | 290 | 280 | 270 |
| 65 | 455 | 444 | 434 | 423 | 413 | 402 | 392 | 381 | 371 | 361 | 350 | 340 | 330 | 320 | 310 | 300 | 290 | 280 | 270 | 260 |
| 64 | 446 | 435 | 424 | 414 | 403 | 392 | 382 | 371 | 361 | 351 | 341 | 330 | 320 | 310 | 300 | 290 | 280 | 270 | 260 | 250 |
| 63 | 436 | 425 | 415 | 404 | 393 | 383 | 372 | 362 | 351 | 341 | 331 | 321 | 310 | 300 | 290 | 280 | 270 | 260 | 250 | 240 |
| 62 | 427 | 416 | 405 | 394 | 384 | 373 | 363 | 352 | 342 | 331 | 321 | 311 | 301 | 290 | 280 | 270 | 260 | 250 | 240 | 230 |
| 61 | 418 | 407 | 396 | 385 | 374 | 364 | 353 | 342 | 332 | 322 | 311 | 301 | 291 | 281 | 270 | 260 | 250 | 240 | 230 | 220 |
| 60 | 408 | 397 | 386 | 375 | 365 | 354 | 343 | 333 | 322 | 312 | 302 | 291 | 281 | 271 | 260 | 250 | 240 | 230 | 220 | 210 |
| 59 | 399 | 388 | 377 | 366 | 355 | 344 | 334 | 323 | 313 | 302 | 292 | 281 | 271 | 261 | 251 | 240 | 230 | 220 | 210 | 200 |
| 58 | 390 | 378 | 367 | 356 | 346 | 335 | 324 | 314 | 303 | 292 | 282 | 272 | 261 | 251 | 241 | 231 | 220 | 210 | 200 | 190 |
| 57 | 380 | 369 | 358 | 347 | 336 | 325 | 315 | 304 | 293 | 283 | 272 | 262 | 252 | 241 | 231 | 221 | 211 | 200 | 190 | 180 |
| 56 | 371 | 360 | 349 | 338 | 327 | 316 | 305 | 294 | 284 | 273 | 263 | 252 | 242 | 231 | 221 | 211 | 201 | 190 | 180 | 170 |
| 55 | 361 | 350 | 339 | 328 | 317 | 306 | 295 | 285 | 274 | 263 | 253 | 242 | 232 | 222 | 211 | 201 | 191 | 181 | 170 | 160 |
| 54 | 352 | 341 | 330 | 319 | 308 | 297 | 286 | 275 | 264 | 254 | 243 | 233 | 222 | 212 | 201 | 191 | 181 | 171 | 161 | 150 |
| 53 | 343 | 331 | 320 | 309 | 298 | 287 | 276 | 265 | 255 | 244 | 233 | 223 | 212 | 202 | 192 | 181 | 171 | 161 | 151 | 140 |
| 52 | 333 | 322 | 311 | 300 | 288 | 277 | 267 | 256 | 245 | 234 | 224 | 213 | 203 | 192 | 182 | 171 | 161 | 151 | 141 | 131 |
| 51 | 324 | 313 | 301 | 290 | 279 | 268 | 257 | 246 | 235 | 225 | 214 | 203 | 193 | 182 | 172 | 162 | 151 | 141 | 131 | 121 |
| 50 | 314 | 303 | 292 | 280 | 269 | 258 | 247 | 236 | 226 | 215 | 204 | 194 | 183 | 173 | 162 | 152 | 141 | 131 | 121 | 111 |
| | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 |

| | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 49 | 305 | 294 | 282 | 271 | 260 | 249 | 238 | 227 | 216 | 205 | 194 | 184 | 173 | 163 | 152 | 142 | 131 | 121 | 111 | 101 |
| 48 | 296 | 284 | 273 | 261 | 250 | 239 | 228 | 217 | 206 | 195 | 185 | 174 | 163 | 153 | 142 | 132 | 122 | 111 | 101 | 091 |
| 47 | 286 | 274 | 263 | 252 | 240 | 229 | 218 | 207 | 196 | 185 | 175 | 164 | 153 | 143 | 132 | 122 | 112 | 101 | 091 | 081 |
| 46 | 276 | 265 | 253 | 242 | 231 | 219 | 208 | 197 | 186 | 176 | 165 | 154 | 144 | 133 | 122 | 112 | 102 | 091 | 081 | 071 |
| 45 | 267 | 255 | 244 | 232 | 221 | 210 | 199 | 188 | 177 | 166 | 155 | 144 | 134 | 123 | 113 | 102 | 092 | 081 | 071 | 061 |
| 44 | 257 | 246 | 234 | 222 | 211 | 200 | 189 | 178 | 167 | 156 | 145 | 134 | 124 | 113 | 103 | 092 | 082 | 071 | 061 | 051 |
| 43 | 248 | 236 | 224 | 213 | 201 | 190 | 179 | 168 | 157 | 146 | 135 | 124 | 114 | 103 | 092 | 082 | 072 | 061 | 051 | 041 |
| 42 | 238 | 226 | 214 | 203 | 191 | 180 | 169 | 158 | 147 | 136 | 125 | 114 | 104 | 093 | 082 | 072 | 062 | 051 | 041 | 031 |
| 41 | 228 | 216 | 205 | 193 | 181 | 170 | 159 | 148 | 137 | 126 | 115 | 104 | 093 | 083 | 072 | 062 | 051 | 041 | 031 | 020 |
| 40 | 218 | 206 | 195 | 183 | 172 | 160 | 149 | 138 | 127 | 116 | 105 | 094 | 083 | 073 | 062 | 052 | 041 | 031 | 021 | 010 |
| 39 | 208 | 196 | 185 | 173 | 161 | 150 | 139 | 128 | 117 | 106 | 095 | 084 | 073 | 063 | 052 | 041 | 031 | 021 | 010 | 0 |
| 38 | 198 | 186 | 175 | 163 | 151 | 140 | 129 | 117 | 106 | 095 | 084 | 074 | 063 | 052 | 042 | 031 | 021 | 010 | 0 | |
| 37 | 188 | 176 | 164 | 153 | 141 | 130 | 118 | 107 | 096 | 085 | 074 | 063 | 053 | 042 | 031 | 021 | 010 | 0 | | |
| 36 | 178 | 166 | 154 | 143 | 131 | 119 | 108 | 097 | 086 | 075 | 064 | 053 | 042 | 032 | 021 | 010 | 0 | | | |
| 35 | 168 | 156 | 144 | 132 | 121 | 109 | 098 | 086 | 075 | 064 | 053 | 043 | 032 | 021 | 011 | 0 | | | | |
| 34 | 158 | 146 | 134 | 122 | 110 | 099 | 087 | 076 | 065 | 054 | 043 | 032 | 021 | 011 | 0 | | | | | |
| 33 | 147 | 135 | 123 | 111 | 100 | 088 | 077 | 065 | 054 | 043 | 032 | 021 | 011 | 0 | | | | | | |
| 32 | 137 | 125 | 113 | 101 | 089 | 078 | 066 | 055 | 044 | 033 | 022 | 011 | 0 | | | | | | | |
| 31 | 126 | 114 | 102 | 090 | 078 | 067 | 055 | 044 | 033 | 022 | 011 | 0 | | | | | | | | |
| 30 | 115 | 103 | 091 | 079 | 068 | 056 | 045 | 033 | 022 | 011 | 0 | | | | | | | | | |
| 29 | 105 | 092 | 080 | 068 | 057 | 045 | 034 | 022 | 011 | 0 | | | | | | | | | | |
| 28 | 094 | 081 | 069 | 057 | 046 | 034 | 023 | 011 | 0 | | | | | | | | | | | |
| 27 | 083 | 070 | 058 | 046 | 034 | 023 | 011 | 0 | | | | | | | | | | | | |
| 26 | 071 | 059 | 047 | 035 | 023 | 011 | 0 | | | | | | | | | | | | | |
| 25 | 060 | 048 | 035 | 023 | 012 | 0 | | | | | | | | | | | | | | |
| 24 | 048 | 036 | 024 | 012 | 0 | | | | | | | | | | | | | | | |
| 23 | 037 | 024 | 012 | 0 | | | | | | | | | | | | | | | | |
| 22 | 025 | 012 | 0 | | | | | | | | | | | | | | | | | |
| 21 | 012 | 0 | | | | | | | | | | | | | | | | | | |
| 20 | 0 | | | | | | | | | | | | | | | | | | | |

| | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

| | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 655 | 647 | 639 | 631 | 624 | 616 | 608 | 600 | 593 | 585 | 577 | 570 | 562 | 554 | 547 | 539 | 531 | 523 | 516 | 508 |
| 99 | 641 | 633 | 625 | 617 | 609 | 601 | 593 | 586 | 578 | 570 | 562 | 554 | 546 | 539 | 531 | 523 | 515 | 507 | 499 | 491 |
| 98 | 627 | 619 | 611 | 603 | 595 | 587 | 579 | 571 | 563 | 555 | 547 | 539 | 531 | 523 | 515 | 507 | 499 | 491 | 483 | 475 |
| 97 | 614 | 605 | 597 | 589 | 581 | 573 | 565 | 557 | 549 | 541 | 532 | 524 | 516 | 508 | 500 | 492 | 483 | 475 | 467 | 459 |
| 96 | 600 | 592 | 584 | 576 | 567 | 559 | 551 | 543 | 535 | 526 | 518 | 510 | 502 | 493 | 485 | 477 | 468 | 460 | 451 | 443 |
| 95 | 587 | 579 | 570 | 562 | 554 | 546 | 537 | 529 | 521 | 512 | 504 | 496 | 487 | 479 | 470 | 462 | 453 | 445 | 436 | 428 |
| 94 | 574 | 566 | 557 | 549 | 541 | 532 | 524 | 515 | 507 | 498 | 490 | 482 | 473 | 464 | 456 | 447 | 439 | 430 | 421 | 413 |
| 93 | 561 | 553 | 544 | 536 | 527 | 519 | 510 | 502 | 493 | 485 | 476 | 468 | 459 | 450 | 442 | 433 | 424 | 416 | 407 | 398 |
| 92 | 549 | 540 | 532 | 523 | 514 | 506 | 497 | 489 | 480 | 471 | 463 | 454 | 445 | 437 | 428 | 419 | 410 | 402 | 393 | 384 |
| 91 | 536 | 528 | 519 | 510 | 502 | 493 | 484 | 476 | 467 | 458 | 450 | 441 | 432 | 423 | 414 | 405 | 397 | 388 | 379 | 370 |
| 90 | 524 | 515 | 507 | 498 | 489 | 480 | 472 | 463 | 454 | 445 | 436 | 428 | 419 | 410 | 401 | 392 | 383 | 374 | 365 | 356 |
| 89 | 512 | 503 | 494 | 486 | 477 | 468 | 459 | 450 | 441 | 432 | 424 | 415 | 406 | 397 | 388 | 379 | 370 | 360 | 351 | 342 |
| 88 | 500 | 491 | 482 | 473 | 464 | 456 | 447 | 438 | 429 | 420 | 411 | 402 | 393 | 384 | 375 | 366 | 356 | 347 | 338 | 329 |
| 87 | 488 | 479 | 470 | 461 | 452 | 443 | 434 | 425 | 416 | 407 | 398 | 389 | 380 | 371 | 362 | 353 | 343 | 334 | 325 | 315 |
| 86 | 476 | 467 | 458 | 449 | 440 | 431 | 422 | 413 | 404 | 395 | 386 | 377 | 368 | 358 | 349 | 340 | 331 | 321 | 312 | 302 |
| 85 | 465 | 456 | 447 | 438 | 428 | 419 | 410 | 401 | 392 | 383 | 374 | 364 | 355 | 346 | 337 | 327 | 318 | 309 | 299 | 290 |
| 84 | 453 | 444 | 435 | 426 | 417 | 408 | 398 | 389 | 380 | 371 | 362 | 352 | 343 | 334 | 324 | 315 | 306 | 296 | 286 | 277 |
| 83 | 442 | 433 | 423 | 414 | 405 | 396 | 387 | 377 | 368 | 359 | 350 | 340 | 331 | 322 | 312 | 303 | 293 | 284 | 274 | 264 |
| 82 | 431 | 421 | 412 | 403 | 394 | 384 | 375 | 366 | 356 | 347 | 338 | 328 | 319 | 310 | 300 | 291 | 281 | 271 | 262 | 252 |
| 81 | 419 | 410 | 401 | 391 | 382 | 373 | 364 | 354 | 345 | 335 | 326 | 317 | 307 | 298 | 288 | 279 | 269 | 259 | 250 | 240 |
| 80 | 408 | 399 | 390 | 380 | 371 | 361 | 352 | 343 | 333 | 324 | 314 | 305 | 296 | 286 | 276 | 267 | 257 | 248 | 238 | 228 |
| 79 | 397 | 388 | 378 | 369 | 360 | 350 | 341 | 331 | 322 | 313 | 303 | 294 | 284 | 274 | 265 | 255 | 246 | 236 | 226 | 216 |
| 78 | 386 | 377 | 367 | 358 | 349 | 339 | 330 | 320 | 311 | 301 | 292 | 282 | 273 | 263 | 253 | 244 | 234 | 224 | 214 | 205 |
| 77 | 375 | 366 | 356 | 347 | 338 | 328 | 319 | 309 | 300 | 290 | 280 | 271 | 261 | 252 | 242 | 232 | 222 | 213 | 203 | 193 |
| 76 | 365 | 355 | 346 | 336 | 327 | 317 | 308 | 298 | 288 | 279 | 269 | 260 | 250 | 240 | 231 | 221 | 211 | 201 | 191 | 181 |
| 75 | 354 | 344 | 335 | 325 | 316 | 306 | 297 | 287 | 277 | 268 | 258 | 249 | 239 | 229 | 219 | 210 | 200 | 190 | 180 | 170 |
| 74 | 343 | 334 | 324 | 315 | 305 | 295 | 286 | 276 | 267 | 257 | 247 | 238 | 228 | 218 | 208 | 199 | 189 | 179 | 169 | 159 |
| 73 | 333 | 323 | 314 | 304 | 294 | 285 | 275 | 265 | 256 | 246 | 236 | 227 | 217 | 207 | 197 | 188 | 178 | 168 | 158 | 148 |
| 72 | 322 | 313 | 303 | 293 | 284 | 274 | 264 | 255 | 245 | 235 | 226 | 216 | 206 | 196 | 186 | 177 | 167 | 157 | 147 | 137 |
| 71 | 312 | 302 | 292 | 283 | 273 | 263 | 254 | 244 | 234 | 225 | 215 | 205 | 195 | 185 | 176 | 166 | 156 | 146 | 136 | 126 |
| 70 | 302 | 292 | 282 | 272 | 263 | 253 | 243 | 233 | 224 | 214 | 204 | 194 | 185 | 175 | 165 | 155 | 145 | 135 | 125 | 115 |
| 69 | 291 | 281 | 272 | 262 | 252 | 242 | 233 | 223 | 213 | 203 | 194 | 184 | 174 | 164 | 154 | 144 | 134 | 124 | 114 | 104 |
| 68 | 281 | 271 | 261 | 252 | 242 | 232 | 222 | 212 | 203 | 193 | 183 | 173 | 163 | 153 | 144 | 134 | 124 | 114 | 104 | 093 |
| 67 | 271 | 261 | 251 | 241 | 231 | 222 | 212 | 202 | 192 | 182 | 173 | 163 | 153 | 143 | 133 | 123 | 113 | 103 | 093 | 083 |
| 66 | 260 | 251 | 241 | 231 | 221 | 211 | 201 | 192 | 182 | 172 | 162 | 152 | 142 | 132 | 122 | 113 | 103 | 092 | 082 | 072 |
| 65 | 250 | 240 | 231 | 221 | 211 | 201 | 191 | 181 | 171 | 162 | 152 | 142 | 132 | 122 | 112 | 102 | 092 | 082 | 072 | 062 |
| 64 | 240 | 230 | 220 | 211 | 201 | 191 | 181 | 171 | 161 | 151 | 141 | 131 | 122 | 112 | 102 | 092 | 082 | 072 | 062 | 051 |
| 63 | 230 | 220 | 210 | 200 | 190 | 181 | 171 | 161 | 151 | 141 | 131 | 121 | 111 | 101 | 091 | 081 | 071 | 061 | 051 | 041 |
| 62 | 220 | 210 | 200 | 190 | 180 | 170 | 161 | 151 | 141 | 131 | 121 | 111 | 101 | 091 | 081 | 071 | 061 | 051 | 041 | 031 |
| 61 | 210 | 200 | 190 | 180 | 170 | 160 | 150 | 140 | 131 | 121 | 111 | 101 | 091 | 081 | 071 | 061 | 051 | 041 | 031 | 020 |
| 60 | 200 | 190 | 180 | 170 | 160 | 150 | 140 | 130 | 120 | 110 | 100 | 091 | 081 | 071 | 061 | 051 | 041 | 030 | 020 | 010 |
| 59 | 190 | 180 | 170 | 160 | 150 | 140 | 130 | 120 | 110 | 100 | 090 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 |
| 58 | 180 | 170 | 160 | 150 | 140 | 130 | 120 | 110 | 100 | 090 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | |
| 57 | 170 | 160 | 150 | 140 | 130 | 120 | 110 | 100 | 090 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | |
| 56 | 160 | 150 | 140 | 130 | 120 | 110 | 100 | 090 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | | |
| 55 | 150 | 140 | 130 | 120 | 110 | 100 | 090 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | | | |
| 54 | 140 | 130 | 120 | 110 | 100 | 090 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | | | | |
| 53 | 130 | 120 | 110 | 100 | 090 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | | | | | |
| 52 | 120 | 110 | 100 | 090 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | | | | | | |
| 51 | 110 | 100 | 090 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | | | | | | | |
| 50 | 100 | 090 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | | | | | | | | |
| | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 |

| | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 49 | 991 | 080 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | | | | | | |
| 48 | 081 | 070 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | | | | | | | |
| 47 | 071 | 060 | 050 | 040 | 030 | 020 | 010 | 0 | | | | | | | | | |
| 46 | 061 | 050 | 040 | 030 | 020 | 010 | 0 | | | | | | | | | | |
| 45 | 051 | 040 | 030 | 020 | 010 | 0 | | | | | | | | | | | |
| 44 | 041 | 030 | 020 | 010 | 0 | | | | | | | | | | | | |
| 43 | 030 | 020 | 010 | 0 | | | | | | | | | | | | | |
| 42 | 020 | 010 | 0 | | | | | | | | | | | | | | |
| 41 | 010 | 0 | | | | | | | | | | | | | | | |
| 40 | 0 | | | | | | | | | | | | | | | | |

| | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

| | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 500 | 492 | 484 | 476 | 469 | 461 | 453 | 445 | 436 | 428 | 420 | 412 | 403 | 395 | 387 | 378 | 369 | 360 | 352 | 343 |
| 99 | 483 | 475 | 467 | 459 | 451 | 442 | 434 | 426 | 418 | 409 | 401 | 392 | 383 | 375 | 366 | 357 | 348 | 339 | 329 | 320 |
| 98 | 466 | 458 | 450 | 442 | 433 | 425 | 416 | 408 | 399 | 391 | 382 | 373 | 364 | 355 | 346 | 337 | 327 | 317 | 308 | 298 |
| 97 | 450 | 442 | 433 | 425 | 416 | 408 | 399 | 390 | 382 | 373 | 364 | 355 | 345 | 336 | 327 | 317 | 307 | 297 | 287 | 277 |
| 96 | 435 | 426 | 417 | 409 | 400 | 391 | 382 | 373 | 364 | 355 | 346 | 337 | 327 | 318 | 308 | 298 | 288 | 278 | 268 | 257 |
| 95 | 419 | 410 | 402 | 393 | 384 | 375 | 366 | 357 | 348 | 338 | 329 | 319 | 310 | 300 | 290 | 280 | 270 | 259 | 249 | 238 |
| 94 | 404 | 395 | 386 | 377 | 368 | 359 | 350 | 341 | 331 | 322 | 312 | 303 | 293 | 283 | 273 | 262 | 252 | 241 | 231 | 219 |
| 93 | 389 | 380 | 371 | 362 | 353 | 344 | 334 | 325 | 315 | 306 | 296 | 286 | 276 | 266 | 256 | 245 | 235 | 224 | 213 | 202 |
| 92 | 375 | 366 | 356 | 347 | 338 | 329 | 319 | 310 | 300 | 290 | 280 | 270 | 260 | 250 | 240 | 229 | 218 | 207 | 196 | 185 |
| 91 | 360 | 351 | 342 | 333 | 323 | 314 | 304 | 295 | 285 | 275 | 265 | 255 | 245 | 234 | 224 | 213 | 202 | 191 | 180 | 168 |
| 90 | 346 | 337 | 328 | 318 | 309 | 299 | 290 | 280 | 270 | 260 | 250 | 240 | 229 | 219 | 208 | 197 | 186 | 175 | 164 | 152 |
| 89 | 333 | 323 | 314 | 304 | 295 | 285 | 275 | 266 | 256 | 246 | 235 | 225 | 215 | 204 | 193 | 182 | 171 | 160 | 148 | 136 |
| 88 | 319 | 310 | 300 | 291 | 281 | 271 | 261 | 251 | 241 | 231 | 221 | 211 | 200 | 189 | 178 | 167 | 156 | 145 | 133 | 121 |
| 87 | 306 | 296 | 287 | 277 | 267 | 258 | 248 | 238 | 228 | 217 | 207 | 196 | 186 | 175 | 164 | 153 | 142 | 130 | 118 | 106 |
| 86 | 293 | 283 | 274 | 264 | 254 | 244 | 234 | 224 | 214 | 204 | 193 | 183 | 172 | 161 | 150 | 139 | 127 | 116 | 104 | 092 |
| 85 | 280 | 270 | 261 | 251 | 241 | 231 | 221 | 211 | 200 | 190 | 180 | 169 | 158 | 147 | 136 | 125 | 114 | 102 | 090 | 078 |
| 84 | 267 | 258 | 248 | 238 | 228 | 218 | 208 | 198 | 187 | 177 | 166 | 156 | 145 | 134 | 123 | 111 | 100 | 088 | 076 | 064 |
| 83 | 255 | 245 | 235 | 225 | 215 | 205 | 195 | 185 | 174 | 164 | 153 | 143 | 132 | 121 | 110 | 098 | 087 | 075 | 063 | 051 |
| 82 | 242 | 233 | 223 | 213 | 203 | 193 | 182 | 172 | 162 | 151 | 140 | 130 | 119 | 108 | 097 | 085 | 074 | 062 | 050 | 038 |
| 81 | 230 | 220 | 210 | 200 | 190 | 180 | 170 | 160 | 149 | 139 | 128 | 117 | 106 | 095 | 084 | 072 | 061 | 049 | 037 | 025 |
| 80 | 218 | 208 | 198 | 188 | 178 | 168 | 158 | 147 | 137 | 126 | 115 | 105 | 094 | 083 | 071 | 060 | 048 | 037 | 025 | 012 |
| 79 | 206 | 196 | 186 | 176 | 166 | 156 | 146 | 135 | 125 | 114 | 103 | 092 | 081 | 070 | 059 | 048 | 036 | 024 | 012 | 0 |
| 78 | 195 | 185 | 175 | 164 | 154 | 144 | 134 | 123 | 113 | 102 | 091 | 080 | 069 | 058 | 047 | 035 | 024 | 012 | 0 | |
| 77 | 183 | 173 | 163 | 153 | 143 | 132 | 122 | 111 | 101 | 090 | 079 | 068 | 057 | 046 | 035 | 023 | 012 | 0 | | |
| 76 | 172 | 161 | 151 | 141 | 131 | 121 | 110 | 100 | 089 | 078 | 068 | 057 | 046 | 034 | 023 | 012 | 0 | | | |
| 75 | 160 | 150 | 140 | 130 | 119 | 109 | 099 | 088 | 078 | 067 | 056 | 045 | 034 | 023 | 011 | 0 | | | | |
| 74 | 149 | 139 | 129 | 118 | 108 | 098 | 087 | 077 | 066 | 055 | 045 | 034 | 023 | 011 | 0 | | | | | |
| 73 | 138 | 128 | 117 | 107 | 097 | 086 | 076 | 065 | 055 | 044 | 033 | 022 | 011 | 0 | | | | | | |
| 72 | 127 | 117 | 106 | 096 | 086 | 075 | 065 | 054 | 044 | 033 | 022 | 011 | 0 | | | | | | | |
| 71 | 116 | 106 | 095 | 085 | 075 | 064 | 054 | 043 | 033 | 022 | 011 | 0 | | | | | | | | |
| 70 | 105 | 095 | 084 | 074 | 064 | 053 | 043 | 032 | 022 | 011 | 0 | | | | | | | | | |
| 69 | 094 | 084 | 074 | 063 | 053 | 043 | 032 | 021 | 011 | 0 | | | | | | | | | | |
| 68 | 083 | 073 | 063 | 053 | 042 | 032 | 021 | 011 | 0 | | | | | | | | | | | |
| 67 | 073 | 063 | 052 | 042 | 032 | 021 | 011 | 0 | | | | | | | | | | | | |
| 66 | 062 | 052 | 042 | 031 | 021 | 011 | 0 | | | | | | | | | | | | | |
| 65 | 052 | 041 | 031 | 021 | 010 | 0 | | | | | | | | | | | | | | |
| 64 | 041 | 031 | 021 | 010 | 0 | | | | | | | | | | | | | | | |
| 63 | 031 | 021 | 010 | 0 | | | | | | | | | | | | | | | | |
| 62 | 021 | 010 | 0 | | | | | | | | | | | | | | | | | |
| 61 | 010 | 0 | | | | | | | | | | | | | | | | | | |
| 60 | 0 | | | | | | | | | | | | | | | | | | | |

| | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

| | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 335 | 324 | 314 | 305 | 295 | 285 | 274 | 264 | 253 | 241 | 229 | 217 | 204 | 190 | 176 | 160 | 143 | 123 | 101 | 071 | 0 |
| 99 | 310 | 300 | 290 | 280 | 269 | 258 | 247 | 235 | 223 | 211 | 197 | 184 | 169 | 153 | 136 | 117 | 096 | 071 | 041 | 0 | |
| 98 | 288 | 277 | 267 | 256 | 245 | 233 | 221 | 209 | 196 | 183 | 168 | 154 | 138 | 121 | 102 | 082 | 059 | 032 | 0 | | |
| 97 | 266 | 256 | 245 | 233 | 222 | 210 | 197 | 184 | 171 | 157 | 142 | 126 | 110 | 092 | 072 | 051 | 027 | 0 | | | |
| 96 | 246 | 235 | 224 | 212 | 200 | 188 | 175 | 161 | 147 | 133 | 118 | 101 | 084 | 066 | 046 | 024 | 0 | | | | |
| 95 | 227 | 215 | 204 | 192 | 179 | 167 | 153 | 140 | 125 | 111 | 095 | 078 | 061 | 042 | 022 | 0 | | | | | |
| 94 | 208 | 197 | 185 | 172 | 160 | 147 | 133 | 119 | 105 | 090 | 074 | 057 | 039 | 020 | 0 | | | | | | |
| 93 | 190 | 178 | 166 | 154 | 141 | 128 | 114 | 100 | 085 | 070 | 054 | 037 | 019 | 0 | | | | | | | |
| 92 | 173 | 161 | 149 | 136 | 123 | 110 | 096 | 082 | 067 | 051 | 035 | 018 | 0 | | | | | | | | |
| 91 | 156 | 144 | 132 | 119 | 106 | 092 | 078 | 064 | 049 | 033 | 017 | 0 | | | | | | | | | |
| 90 | 140 | 128 | 115 | 102 | 089 | 076 | 062 | 047 | 032 | 016 | 0 | | | | | | | | | | |
| 89 | 124 | 112 | 099 | 086 | 073 | 059 | 045 | 031 | 016 | 0 | | | | | | | | | | | |
| 88 | 109 | 097 | 084 | 071 | 058 | 044 | 030 | 015 | 0 | | | | | | | | | | | | |
| 87 | 094 | 082 | 069 | 056 | 043 | 029 | 015 | 0 | | | | | | | | | | | | | |
| 86 | 080 | 067 | 055 | 041 | 028 | 014 | 0 | | | | | | | | | | | | | | |
| 85 | 066 | 053 | 040 | 027 | 014 | 0 | | | | | | | | | | | | | | | |
| 84 | 052 | 039 | 027 | 013 | 0 | | | | | | | | | | | | | | | | |
| 83 | 039 | 026 | 013 | 0 | | | | | | | | | | | | | | | | | |
| 82 | 025 | 013 | 0 | | | | | | | | | | | | | | | | | | |
| 81 | 013 | 0 | | | | | | | | | | | | | | | | | | | |
| 80 | 0 | | | | | | | | | | | | | | | | | | | | |

| | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

PSYCHOMETRIKA—VOL. 12, NO. 1
MARCH, 1947

# THE USE OF PSYCHOLOGICAL TECHNIQUES IN MEASURING AND CRITICALLY ANALYZING NAVIGATORS' FLIGHT PERFORMANCE

LAUNOR F. CARTER
UNIVERSITY OF ROCHESTER
AND
FRANK J. DUDEK*
UNIVERSITY OF SOUTHERN CALIFORNIA

Under controlled flight conditions, the distance between a navigator's report of position and his actual position is a criterion of success in dead reckoning navigation. Students' logs were evaluated for five separate missions by comparing the students' entries with standards determined by experts. The reliability of this technique is indicated by the fact that mission to mission intercorrelations of error scores were low, while the intercorrelations between legs of the same mission were moderately high. The intercorrelations between the error scores for the different navigation variables were computed and analyzed by using both factor analysis and multiple regression techniques. Both analyses indicated that a major portion of all dead reckoning error could be attributed to errors made in determining magnetic deviation. As a result of these analyses, recommendations were made for changing the instruction in dead reckoning and alterations in the equipment used were suggested.

## 1. *The Problem*

Since one of the most difficult tasks in psychology is the objective evaluation of complex skills, one of the most difficult tasks of the aviation psychologist is the objective evaluation of flight performance. It is the purpose of this paper to discuss a method for evaluating navigators' dead reckoning proficiency as an example of the difficulties encountered in assessing aerial performance and also to describe a method by which a successful evaluative technique may be used to analyze performance critically and to lead to its improvement.

At first glance it would seem that the evaluation of dead reckoning would be simple and straightforward. The navigator's plane departs from a known position; by determining the true air speed, the wind, and the true course flown, the navigator calculates his position for a particular time without recourse to check points on the ground. By comparing the navigator's calculated position with his actual position, it is possible to obtain an immediate, objective measure of the accuracy of his work. In actual practice this can reasonably be done

* This work was done as a part of the AAF Aviation Psychology Program. The authors are indebted to Thomas Paltier, Harley Smith, Wolcott, Lyon, and John King for assistance with the calculations.

for the navigators in any one plane, but as soon as it is desired to compare the performance of navigators from plane to plane, it is found that the results are not comparable due to a multitude of uncontrolled flight conditions. Even though two planes depart from the same point for the same destination, it is not proper to evaluate the navigators' relative skill on the basis of their objective results in terms of position error unless most careful controls have been instituted to assure similar flight conditions for both planes. The development of an objective scale for evaluating navigation skill involved setting up conditions in which all the navigators would be exposed to the same situation in the air, perform the same kind of navigation, and have their performances evaluated in terms of a precise standard and in an objective and equitable manner. At the same time, the technique used had to insure that the missions were flown safely, represented an efficient expenditure of plane time, and were consistent with the other operational problems encountered at a training school. In addition, an adequate technique had to be reliable and reproducible.

### 2. *Description of the Technique*

The technique employed was to fly the navigation missions in formation and to require the navigators to perform "follow-the-pilot" navigation, that is, dead reckoning navigation in which the navigator simply determines the position of the plane at specified times but does not direct the pilot regarding the course he is to fly. By flying formations in which the planes were within a few yards of each other, all the students were flown over the same course, at the same speed, and under the same weather conditions. Several different types of formations and planes were used during this experiment.

Each mission consisted of four legs with each leg covering approximately 100 miles and terminating at selected turning points. At each turning point the formation changed course by approximately 90°. The turns were made consistently in one direction so that a somewhat square track was followed. Flying the missions over a square track yielded three advantages: (1) it was possible to return to the place of departure and thus secure maximal use of student and plane time; (2) by turning 90° at each turning point, the students were provided with an opportunity to obtain the best estimate of wind on two headings; and (3) four legs of 100 miles each made possible the collection of adequate, comparable, and independent samples of navigation on each leg. At the same time that the students were required to determine their position at each of the four turning points by dead reckoning, expert lead navigators in the first plane of the formation were directing the course flown and keeping a precise record of the actual

track made good over the ground. While the student navigators had to rely only on their calculations, the lead navigators actually looked at the ground and carefully determined where they were at all times. Other personnel in the planes made very accurate readings of the air speed meter, the compass, the drift meter, and the astro-compass. In this way extremely accurate standards were developed with which the students' log entries could be compared. Since the students all navigated under the same flight conditions and objective standards were available for evaluating their performance, it was believed that an adequate technique had been developed.

### 3. *The Reliability of the Technique*

In October and November of 1944, approximately 180 students were flown on three different flight missions, and again during the summer of 1945, 80 students were flown on four different flight missions spaced throughout the navigation course. From the data collected on these missions, it was possible to analyze the reliability of the technique. Two types of analysis were possible, namely, an intra-mission reliability which corresponds to a split-half reliability and an inter-mission reliability which corresponds to a test-retest reliability.

Since the missions were flown by legs, there were four independent evaluations of each navigator's performance. The correlation between the errors made on the first and fourth legs and the errors made on the second and third legs were computed for a number of navigation variables; however, only the reliabilities for true air speed

TABLE 1

Within Missions Reliability Coefficients for True Air Speed (Legs 1-4 With Legs 2-3)

| Flight | Formation | Mission 5 | | Mission 6 | | Mission 7 | | All Missions |
|---|---|---|---|---|---|---|---|---|
| | | *N* | Rho | *N* | Rho | *N* | Rho | Combined *r* |
| 23A | 1 | 23 | .88 | 23 | .56 | 22 | .76 | .78 |
| | 2 | 19 | .25 | 18 | .75 | 17 | .86 | .58 |
| 23B | 1 | 23 | .59 | 23 | .34 | 20 | .21 | .48 |
| | 2 | 20 | .25 | 20 | .72 | 20 | .61 | .57 |
| 24A | 1 | 21 | .51 | 21 | .68 | 24 | .75 | .68 |
| | 2 | 17 | .94 | 17 | .57 | 21 | .27 | .70 |
| 24B | 1 | 21 | .56 | 21 | .64 | 24 | .63 | .63 |
| | 2 | 17 | .45 | 18 | .33 | 20 | .43 | .42 |
| Combined *r* | | | .61 | | .61 | | .63 | $r'_{11} = .62$ |
| | | | | | | | | $r_{11} = .77$ |

will be discussed at this point since the conclusions drawn from the true air speed reliabilities are similar to those drawn from the remainder of the data. Table 1 shows the correlation between the error scores of the first and fourth legs and those of the second and third legs for the first group of students.

Since the error scores obtained were not normally distributed and it was felt that extreme scores should not bias the reliabilities, rank order correlations rather than Pearsonian correlations were calculated. However, the combined correlations were obtained by converting the rho's to $r$'s and combining them by the $z$ transformation and reconverting to $r$'s. From Table 1 it will be seen that while there is considerable variability from formation to formation, the reliabilities are fairly high, the Spearman-Brown correction giving a reliability of .77. But from examining the data, it was suspected that this reliability was unduly high due to constant factors within particular planes; thus, if the air speed meter in one plane were improperly calibrated, all of the students within that plane would make high error scores which would not be indicative of their true ability. If there were no systematic factors within planes, it would be expected that the correlation between the error scores of students within the same planes would be 0. Table 2 shows the correlation of the error scores of students in the same planes by seats. In view of the magnitude of these correlations, it was thought probable that systematic plane differences accounted for some of the reliability shown in Table 1.

TABLE 2
Correlation Coefficients
Seats for True Air Speed on Mission 5

| Flight | Seats 1 and 2 | | Seats 2 and 3 | | Seats 1 and 3 | |
|---|---|---|---|---|---|---|
| | $N$ | Rho | $N$ | Rho | $N$ | Rho |
| 23A | 14 | .36 | 13 | .54 | 13 | .22 |
| 23B | 14 | .20 | 13 | —.12 | 14 | —.11 |
| 24A | 10 | .10 | 10 | .49 | 11 | .34 |
| 24B | 11 | .21 | 12 | .24 | 10 | .65 |
| Combined $r$ | | .25 | | .30 | | .28 |

These results indicated that a more appropriate reliability might be obtained by analysis of covariance. After the variances associated with plane and seat were removed, a coefficient of .48 was obtained for the ultra-mission reliability of the individual's scores.

Another estimate of the technique's reliability was obtained by

comparing performance on successive missions. The error scores between missions were computed and the results are shown in Table 3.

TABLE 3
Between Missions Correlation Coefficients for True Air Speed

| Flight | For- | Mission 5 vs. 6 | | Mission 6 vs. 7 | | Mission 5 vs. 7 | |
|---|---|---|---|---|---|---|---|
| | mation | *N* | Rho | *N* | Rho | *N* | Rho |
| 23A | 1 | 23 | —.13 | 22 | .04 | 22 | —.08 |
| | 2 | 18 | .03 | 17 | —.27 | 17 | —.16 |
| 23B | 1 | 23 | —.08 | 20 | —.35 | 20 | —.06 |
| | 2 | 20 | —.05 | 20 | .00 | 20 | .28 |
| 24A | 1 | 19 | —.61 | 21 | .21 | 21 | .14 |
| | 2 | 15 | .17 | 17 | .13 | 17 | —.19 |
| 24B | 1 | 18 | .45 | 21 | —.10 | 21 | .01 |
| | 2 | 15 | —.34 | 18 | .08 | 17 | —.11 |
| Combined *r* | | | —.01 | | .00 | | .00 |

This complete lack of relationship raises serious question regarding the adequacy of the technique. However, since the experiments took place at a time when the learning curve was steep, the test-retest reliability coefficient may have been somewhat attenuated. There were probably other uncontrolled variables, such as day-to-day fluctuations in the planes, the relative skill of pilots, and variable weather conditions. It is, of course, also possible that the technique of measurement is sufficiently reliable, but that the navigation task is so difficult and so influenced by chance factors that consistent test-retest results could not be expected except for a great number of missions. If this is the case, it raises serious doubt as to the possibility of ever practically achieving an adequate objective evaluation of navigators' performance. It should be remembered that each student navigated over twelve hours in furnishing the data on which these reliabilities are based.

Since these reliabilities are based on data collected from the first application of the technique and since somewhat unstable AT-7 planes seating only three students were used, it was decided to apply the technique to a second sample of 80 cadets at Ellington Field where larger and more stable C-47's, seating eight navigators, were available. The Ellington Field students were flown on missions in their seventh, eleventh, sixteenth, and twentieth weeks of training. Table 4 shows the correlation between successive missions and between the total errors

on missions 1 and 4 against those on missions 2 and 3. In addition to correlations for true air speed, the correlations for deviation, drift, and distance-off are shown.

TABLE 4

Between Missions Correlation Coefficients from the Second Series of Missions

| Variable | Missions 1 and 2 | Missions 2 and 3 | Missions 3 and 4 | Missions 1 + 4 and 2 + 3 |
|---|---|---|---|---|
| Drift | .22 | .22 | —.01 | .19 |
| Deviation | .11 | .13 | —.03 | .09 |
| True Air Speed | .27 | .18 | .13 | .27 |
| Distance-Off | —.10 | .13 | .01 | —.03 |

Certainly, the intercorrelations in this table are not high enough to suggest that this technique, even though internally reliable, will show consistent positive correlations when administered from time to time.

Similar correlations have been reported for other measures of complex flying skill. The following data have been reported by Psychological Research Project (Pilot), Randolph Field, Texas, in their *Report for the Fiscal Year 1945*. Table 5 shows the reliability for data collected in measuring the performance of elementary pilot students in landing.

TABLE 5

Reliability of Measures of Elementary Pilot Students' Landing Ability

| Measures | Ground vs. Air Observer | | 1st Landing vs. 2nd Landing on | | 1st Day Landings vs. 2nd Day Landings | |
|---|---|---|---|---|---|---|
| | Trial 1 | Trial 2 | Day 1 | Day 2 | Landing 1 | Landing 2 |
| Zone in which Plane Lands | .79* | .86 | .14† | .11 | .02† | .01 |
| Landing Attitude | .87 | .68 | .45 | .18 | —.07 | .15 |
| Bounced or Dropped | .89 | .87 | .44 | .19 | —.01 | .01 |

* The $N$ for these six correlations equals 152.
† The $N$ for these six correlations equals 170.

Again it will be noted that the reliability of ratings made for any particular mission is high while the reliability of ratings made for different missions on different days tends to be low. Without offering further data, it may be ventured that most measures of flight performance for the navigator, pilot, gunner, and bombardier will tend to

show low inter-mission correlations even though intra-mission reliabilities may be fairly high.

4. *Use of the Technique in Critically Analyzing Performance*

Whenever it is possible to measure the component parts of any performance, an objective assessment of the relative importance of the several kinds of operator errors may be undertaken. On logical grounds it was possible to attribute poor performance to a number of different types of error in navigation; but it was not possible to determine objectively the relative importance of these errors until some technique had been developed to measure the accuracy both of over-all performance and of each of the steps of navigation. The technique previously described made it possible to determine the particular operations responsible for most of the navigator's error. By correlating the error scores for each navigation operation, a correlation matrix for the error scores was constructed. In analyzing this matrix, either by factor analysis techniques or by multiple regression techniques, it was possible to determine the major causes of navigation error and their relative importance. Tables 6 and 7 show the correlations between the error scores for the different navigation variables for three sets of data.

TABLE 6
Intercorrelations Between Errors for 8 Navigation Variables for 165 Students Who Were in Their 7th Week of Training at Selman Field

| | Track | Drift | DEV | TAS | WF | WD | GS | DO | MEAN | S.D. |
|---|---|---|---|---|---|---|---|---|---|---|
| Track | | .11 | .76 | −.03 | .12 | .17 | .06 | .78 | 13.40 | 8.04 |
| Drift | .11 | | .01 | .00 | .72 | .27 | .52 | .12 | 5.10 | 3.35 |
| Deviation | .76 | .01 | | −.02 | .07 | .17 | .11 | .67 | 13.30 | 8.10 |
| True Air Speed | −.03 | .00 | −.02 | | .04 | −.13 | .28 | .12 | 8.20 | 6.65 |
| Wind Force | .12 | .72 | .07 | .04 | | .12 | .57 | .18 | 10.30 | 8.10 |
| Wind Direction | .17 | .27 | .17 | −.13 | .12 | | .27 | .18 | 135.00 | 87.00 |
| Ground Speed | .06 | .52 | .11 | .28 | .57 | .27 | | .30 | 19.20 | 9.76 |
| Distance-Off | .78 | .12 | .67 | .12 | .18 | .18 | .30 | | 31.40 | 14.80 |

It should be noted that Table 7 shows two sets of data, one set above the diagonal and the other below. Table 8 shows the rotated factor loadings and communalities for the three sets of correlations shown in Tables 6 and 7.

TABLE 7

Intercorrelation Between Errors for 14 Navigation Variables for Two Groups* of 40 Students Each in Their 7th Week of Training

| | | FLIGHTS A & B | | | | | | | | | | | | | A — B | | C — D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Track | Drift | TH | MH | DEV | IAS | CAS | TAS | WF | WD | GS | DIST | Time | DO | MEAN | S.D. | MEAN | S.D. |
| Flights C & D | Track (°) | | .07 | .89 | .80 | .77 | −.13 | −.07 | −.13 | −.11 | .04 | −.06 | .00 | .17 | .68 | 21.70 | 9.49 | 26.05 | 12.36 |
| | Drift (°) | .16 | | .04 | −.02 | .04 | .03 | .07 | .10 | .03 | .80 | .51 | .46 | .43 | .41 | 5.95 | 4.79 | 10.30 | 4.29 |
| | True Heading (°) | .82 | .27 | | .93 | .92 | .03 | .07 | .08 | .03 | .05 | −.07 | .05 | .14 | .68 | 20.83 | 8.06 | 25.08 | 9.23 |
| | Magnetic Heading (°) | .83 | .13 | .90 | | .95 | .09 | .01 | .16 | .07 | .03 | −.03 | .03 | .14 | .67 | 20.68 | 8.16 | 24.26 | 9.36 |
| | Deviation (°) | .83 | .13 | .90 | 1.00 | | .18 | .23 | .18 | .00 | .12 | .01 | .09 | .14 | .72 | 20.84 | 7.82 | 24.26 | 9.36 |
| | Indicated Air Speed (K) | −.21 | −.18 | −.24 | −.21 | −.21 | | .96 | .63 | .08 | .13 | .48 | .30 | .06 | .13 | 7.39 | 3.30 | 5.24 | 3.49 |
| | Calibrated Air Speed (K) | −.20 | −.05 | −.21 | −.21 | −.22 | .90 | | .69 | .06 | .15 | .55 | .28 | .08 | .20 | 6.58 | 2.86 | 4.63 | 2.76 |
| | True Air Speed (K) | −.31 | −.23 | −.40 | −.37 | −.37 | .83 | .76 | | .28 | −.02 | .38 | .27 | .14 | .09 | 6.30 | 2.59 | 5.76 | 3.34 |
| | Wind Force (K) | .25 | .85 | .30 | .20 | .19 | −.18 | −.03 | −.22 | | −.03 | .13 | −.23 | −.22 | .00 | 10.57 | 7.12 | 17.35 | 8.45 |
| | Wind Direction (°) | .24 | .55 | .27 | .20 | .17 | .05 | .17 | −.01 | .54 | | .55 | .41 | .20 | .05 | 86.50 | 70.61 | 42.08 | 21.45 |
| | Ground Speed (K) | .07 | .35 | .09 | .04 | .04 | .25 | .34 | .14 | .47 | .35 | | .48 | .11 | .37 | 23.75 | 13.71 | 23.13 | 13.67 |
| | Distance Traveled (NM) | .00 | .25 | .06 | .02 | .02 | .23 | .33 | .07 | .26 | .33 | .82 | | .58 | .36 | 21.62 | 15.73 | 24.21 | 13.28 |
| | Time Traveled (Min) | −.01 | −.07 | .05 | .08 | .07 | −.06 | −.12 | −.10 | −.09 | .09 | .10 | .50 | | .20 | 6.11 | 3.73 | 4.31 | 1.87 |
| | Distance-Off (NM) | .83 | .21 | .71 | .73 | .73 | −.15 | −.06 | −.25 | .39 | .29 | .34 | .30 | .10 | | 63.30 | 32.36 | 62.40 | 26.75 |

* Note that there are two sets of intercorrelations in this table, one above and the other below the diagonal.

TABLE 8
Rotated Factor Loadings for Three Different Missions

| Navigation Variables | 165 Students in 7th Week at Selman | | | | | 40 Students in 7th Week at Ellington | | | | 40 Other Students in 7th Week at Ellington | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Factors | I | II | III | IV | $h^2$ | I | II | III | $h^2$ | I | II | III | IV | $h^2$ |
| Track | .88 | .15 | .16 | .03 | .82 | .88 | .07 | −.19 | .82 | .88 | .10 | .06 | .02 | .79 |
| Drift | .00 | .90 | .09 | .00 | .82 | .06 | .89 | .04 | .80 | .10 | .89 | −.11 | −.01 | .81 |
| True Heading | * | | | | | .97 | −.02 | .06 | .94 | .92 | .19 | .03 | −.02 | .88 |
| Magnetic Heading | | | | | | .94 | −.06 | .12 | .90 | .98 | .03 | −.14 | .03 | .98 |
| Deviation | .87 | .01 | .00 | .01 | .77 | .94 | −.01 | .21 | .93 | .98 | .03 | −.14 | .03 | .98 |
| Indicated Air Speed | | | | | | −.01 | .01 | .90 | .81 | −.28 | −.08 | .91 | −.01 | .91 |
| Calibrated Air Speed | | | | | | .05 | .13 | .87 | .78 | −.29 | .09 | .91 | .01 | .92 |
| True Air Speed | −.03 | −.06 | .41 | −.16 | .20 | .04 | .04 | .82 | .68 | −.42 | −.12 | .75 | −.09 | .76 |
| Wind Force | .07 | .77 | .29 | −.16 | .71 | −.03 | −.17 | .24 | .09 | .17 | .90 | −.08 | .07 | .85 |
| Wind Direction | .17 | .28 | .02 | .49 | .35 | −.01 | .69 | .12 | .49 | .17 | .62 | .15 | .14 | .46 |
| Ground Speed | .13 | .49 | .76 | .14 | .85 | −.03 | .63 | .47 | .62 | −.04 | .49 | .34 | .57 | .68 |
| Distance Traveled | | | | | | .05 | .72 | .18 | .55 | .00 | .28 | .30 | .91 | .99 |
| Time Traveling | | | | | | .11 | .48 | .01 | .24 | .06 | −.09 | −.08 | .51 | .28 |
| Distance-Off | .79 | .10 | .44 | .02 | .83 | .75 | .37 | .08 | .71 | .78 | .19 | .13 | .31 | .76 |

* The variables for which no loadings appear were not included in this analysis.

These factor loadings were obtained by the centroid method of factor analysis. In these analyses the variable, distance-off, may be considered as a criterion since the types of errors are being sought which contribute most to over-all inaccuracy in dead reckoning navigation. In each analysis in Table 8, the first factor contains the heaviest distance-off loading and thus the variables defining this factor will be those which are responsible for the largest part of dead reckoning error. It will be noted that deviation has a high loading in the first factor in each of the analyses. Since deviation and drift are the two independent variables making up track, it is apparent that errors in deviation are the major cause of error in dead reckoning navigation. (Drift and deviation are the two independent variables which when applied to compass heading determine track; errors in the other "heading variables" such as magnetic heading, true heading, and track are directly attributable to errors in deviation or drift.)

The second factor in each analysis has a high loading in drift and also a high loading in one of the wind variables; but it has only moderate or low factor loadings on the criterion, distance-off. Similarly, the third factor has its highest loadings in one of the speed variables, ground speed or air speed; and this factor also has only moderate or low loadings in the criterion. The fourth factor, in the cases when it has been extracted, may be a specific factor associated with the particular mission involved.

From this analysis two important points stand out. First, in all the analyses the first factor is most clearly attributable to error in the determination of deviation and it is also the most important factor in accounting for navigation inaccuracies. The second point is that in each analysis three factors, each identifiable by the same loadings from analysis to analysis, are clearly identifiable.

It may be suggested that in each analysis there are certain loadings for one of the factors which do not appear in other analyses. This is only to be expected. The psychologist should not think of these missions as similar to carefully controlled testing situations since, as has been previously mentioned, the actual testing situation changed from mission to mission. The wind forces were considerably different on each mission; the type of plane and the type of formation used at Selman Field differed from those used at Ellington Field. Thus it is surprising that the results were as consistent as they are. Another reason for these fluctuations is the small number of cases in the two Ellington Field groups. They were not analyzed together since for the first flight the average wind force was between 10 and 15 knots while for the second flight it was between 20 and 25 knots. The influence of the

number of cases may be noticed in the intercorrelations and factor loadings for drift, wind force, and wind direction. In the first Ellington Field analysis three students obtained reciprocal winds by plotting negative drifts when they should have been positive and vice versa. The result was that there were no errors in wind force to correspond to the large discrepancies between actual drift readings and the students' estimates, but the error in wind direction was at a maximum. Thus, the correlation between errors for wind force and drift was low, while the correlation between errors for wind direction and drift was high. Even in spite of these errors the over-all factorial composition is quite consistent.

To check and extend the above results, the correlation matrices presented in Tables 6 and 7 and the results from three later missions were analyzed by the use of multiple regression techniques. The variable, distance-off, was considered the independent variable and beta weights were determined for the other variables. Thus the highest beta weights would indicate those variables whose errors contributed the most to distance-off in dead reckoning navigation. Table 9 shows the beta weights and multiple correlations obtained when distance-off is predicted by seven navigation variables, and also when it is predicted by the three completely independent and basic variables.

TABLE 9

Beta Weights and Multiple Correlations Obtained for Predicting Distance-Off from Seven Navigational Variables and from the Three Basic Navigational Variables

| | Group | Selman | Ellington | | | |
|---|---|---|---|---|---|---|
| | Week in Training | 7th | 7th | 11th | 16th | 20th |
| Variables | Track | .67 | .53 | .84 | .77 | .61 |
| | Drift | .00 | .09 | .00 | .02 | .00 |
| | Deviation | .13 | .28 | .00 | .20 | .05 |
| | True Air Speed | .08 | .00 | .00 | .24 | .07 |
| | Wind Force | .00 | .00 | .00 | .16 | .00 |
| | Wind Direction | .00 | .00 | .16 | .05 | .07 |
| | Ground Speed | .23 | .31 | .27 | .24 | .39 |
| | $R_{8\cdot1\ldots7}$ | .83 | .87 | .91 | .99 | .85 |
| Variables | Drift | .11 | .25 | .15 | .36 | .09 |
| | Deviation | .67 | .71 | .60 | .76 | .41 |
| | True Air Speed | .13 | .01 | −.02 | .10 | .29 |
| | $R_{4\cdot123}$ | .69 | .77 | .63 | .84 | .49 |

It will be seen that the results of the factor analyses are confirmed and that deviation is again the most important of the variables (remembering that errors in track are a function of errors in deviation). The relative importance of errors in true air speed and in drift seems to be about the same, but both are considerably less important than errors in deviation.

On the basis of these results, it was possible to recommend to the personnel responsible for navigation training that the amount of instruction and practice on the different techniques for determining deviation be increased. The results also indicated that the instruments used for determining deviation might be faulty and the consideration of these instruments led to the recommendation of several changes in the astro-compass to improve the accuracy with which it could be used. It will be noted that in studying a complex skill, it is first necessary that a technique be developed for assessing over-all performance of the skill and also for assessing the separate components determining this performance. Once these measures have been developed, it is possible to determine statistically the importance of the various parts of the task. Such an analysis can then be used as a basis for recommending changes in courses of instruction or as a point of departure for examining the different parts of the task to determine those aspects of performance which can be most profitably improved.

## 5. *Summary*

A technique for assessing dead reckoning navigation performance was developed. The technique consisted of flying a large number of students simultaneously in formation over the same course. The reliability of this technique was found to be fairly high within any one mission, but to be practically zero when measured by the correlation between missions. Considering these reliabilities and others collected in measuring pilot's landing ability, it is concluded that in many complex skills reliability for any particular trial may be high and yet the correlation between trials, which corresponds to test-retest reliability, may be low.

Once the technique for assessing performance had been developed, it was possible to determine the principal cause of error in dead reckoning performance by the application of factor analysis and multiple regression techniques. On the basis of these results, it was possible to make recommendations, based on objective evidence, regarding changes in instruction and improvement of the instruments used in navigation.

PSYCHOMETRIKA—VOL. 12, NO. 1
MARCH, 1947

# ANALYSIS IN TERMS OF FREQUENCIES OF DIFFERENCES

HAROLD A. VOSS
THE PROCTOR AND GAMBLE COMPANY

A technique of analysis utilizing frequencies of differences is described and applied to a hypothetical experiment involving two methods of instruction. A nomograph is provided for computing the chi-square values applicable to the method.

On several occasions, the author has had the problem of comparing the effectiveness of two training methods, or aids, where a number of experimental conditions were involved. In such case, the investigator may employ analysis of variance or the conventional techniques for evaluating the reliability of differences. However, either time pressure or the preliminary nature of the investigation may make it advisable to short cut the lengthy computations involved in these methods. The method to be described is believed to fulfill the need for such a short cut.

Briefly, the method involves tabulation of the paired measures under the varied conditions of the experiment. The measures may be scores of individuals, means, standard deviations, percentages, or correlation coefficients, depending on the measuring instrument and other circumstances of the experiment. A second table is then prepared from the first, the entries being the frequency of differences for each condition favoring the one method over the other. For each experimental condition $\chi^2$ is then computed to determine if the distribution of differences differs significantly from the expected distribution. If the two methods were equally effective, the expected distribution of differences would center about zero with an equal number of differences favoring each method.

The hypothetical data in Table 1 have been prepared to illustrate the method. Two methods of instruction in puzzle solving are compared, demonstration alone and demonstration plus explanation. Six groups, each consisting of a subgroup instructed by one method and a subgroup instructed by the other method, have been selected in random fashion from a population of school children relatively homogeneous in age, intelligence, and school grades. For each group the schedule of instruction and test is the same. On Days 1 and 2 a

TABLE 1
Puzzle Solving Under Conditions of Demonstration and of Demonstration Plus Explanation
(Each entry is the mean time in seconds for a group of 10.)
(Hypothetical data)

| | Group 1 | | Group 2 | | Group 3 | | Group 4 | | Group 5 | | Group 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *N* | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| Reward | No | | Yes | | No | | Yes | | No | | Yes | |
| Puzzle Content | Numbers | | Numbers | | Patterns | | Patterns | | Symbols | | Symbols | |
| Instruction | Dem | Dem Exp | Dem | Dem Exp | Dem | Dem Exp | Dem | Dem Exp | Dem | Dem Exp | Dem | Dem Exp |
| Day 1 | | | | | | | | | | | | |
| Puzzle type | | | | | | | | | | | | |
| 1 step | +18.5 | 17.8 | —18.0 | 18.4 | —19.0 | 19.3 | +18.2 | 17.9 | °20.6 | 20.6 | +22.0 | 19.7 |
| 2 step | —26.3 | 27.1 | +25.1 | 24.7 | +29.1 | 27.2 | +28.1 | 26.6 | —31.3 | 32.2 | +30.1 | 29.2 |
| 3 step | —39.2 | 40.5 | —41.3 | 43.1 | +45.2 | 41.5 | +46.2 | 39.1 | +56.1 | 51.9 | +53.3 | 48.1 |
| Day 2 | | | | | | | | | | | | |
| Puzzle type | | | | | | | | | | | | |
| 1 step | +15.5 | 15.2 | —15.3 | 16.1 | —16.3 | 16.8 | +16.2 | 16.0 | +20.3 | 18.6 | +21.1 | 18.9 |
| 2 step | +19.3 | 18.1 | —17.1 | 18.3 | +27.3 | 21.8 | +24.3 | 20.7 | +27.3 | 26.1 | +25.1 | 22.1 |
| 3 step | —36.1 | 37.4 | +38.2 | 33.1 | +43.1 | 38.9 | +41.4 | 35.9 | +55.4 | 53.4 | +52.3 | 46.7 |
| Day 15 | | | | | | | | | | | | |
| Puzzle type | | | | | | | | | | | | |
| 1 step | —17.5 | 17.6 | —16.0 | 16.5 | —17.1 | 17.3 | +17.0 | 15.5 | +19.1 | 18.2 | +20.0 | 18.6 |
| 2 step | —22.3 | 23.5 | +20.3 | 19.4 | +26.8 | 26.4 | +24.1 | 21.6 | +28.2 | 26.3 | +26.1 | 25.4 |
| 3 step | +37.1 | 35.7 | +36.1 | 33.9 | +38.1 | 35.6 | +41.3 | 34.8 | °51.3 | 51.3 | +51.4 | 46.3 |

period of instruction is followed by a test on three puzzles and two weeks later, on Day 15, there is another test on three puzzles but no instruction. Use of rewards and differences in puzzle content are varied systematically over the groups. To summarize, puzzle solving under two types of instruction, demonstration and demonstration plus explanation, is compared over the following conditions:

Motivation — reward used
reward not used
Test — after first instruction period
after second instruction period
two weeks after second instruction period
Puzzle content — numbers
geometric patterns
symbols other than numbers
Puzzle type — one step
two step
three step

Each pair of entries in Table 1 is marked with a plus (+) if the difference favors the demonstration plus explanation group. Zero (0) indicates no difference, and minus (—) indicates a difference in favor of the demonstration group. Table 2 presents the tabulation of the frequency of + and — differences for each experimental condition. Zero differences give a credit of .5 to each method. For each subcategory of the experimental conditions, chi square has been computed to determine whether the observed distribution of differences departs significantly from the expected distribution. As was pointed out previously, if neither method is superior, an equal number of differences can be expected to favor each method in accordance with the familiar null hypothesis.

In the chi square formula,

$$\chi^2 = \Sigma[(f_o - f_t)^2/f_t] ,$$

where

$f_o$ = observed frequency; the values used in the present case are the frequency of plus and of minus differences in turn;

$f_t$ = expected frequency, the value used is $n/2$.

The analysis in Table 2 shows that demonstration plus explanation is a more effective method of instruction in puzzle solving than

TABLE 2

Analysis in Terms of Frequencies of Differences Favoring One Method of Instruction over the Other

(Based on hypothetical data from Table 1)

| Experimental condition | $n$ | $f_0$ (+) Dem Exp | $f_0$ (—) Dem | $f_t$ ($n/2$) | $\chi^{2*}$ | $P$ $df$=1 |
|---|---|---|---|---|---|---|
| **Motivation** | | | | | | |
| Reward | 27 | 22.0 | 5.0 | 13.5 | 10.70 | <.01 |
| No reward | 27 | 17.0 | 10.0 | 13.5 | 1.81 | .20—.10 |
| **Test** | | | | | | |
| Day 1 | 18 | 11.5 | 6.5 | 9.0 | 1.39 | .30—.20 |
| Day 2 | 18 | 14.0 | 4.0 | 9.0 | 5.56 | .02—.01 |
| Day 15 | 18 | 13.5 | 4.5 | 9.0 | 4.50 | .05—.02 |
| **Puzzle Content** | | | | | | |
| Numbers | 18 | 8.0 | 10.0 | 9.0 | .22 | .70—.50 |
| Patterns | 18 | 15.0 | 3.0 | 9.0 | 8.00 | <.01 |
| Symbols | 18 | 16.0 | 2.0 | 9.0 | 10.89 | <.01 |
| **Puzzle type** | | | | | | |
| 1 step | 18 | 10.5 | 7.5 | 9.0 | .50 | .50—.30 |
| 2 step | 18 | 14.0 | 4.0 | 9.0 | 5.56 | .02—.01 |
| 3 step | 18 | 14.5 | 3.5 | 9.0 | 6.72 | <.01 |
| **Total experiment** | 54 | 39.0 | 15.0 | 27.0 | 10.67 | <.01 |

* $\chi^2 = \Sigma[(f_0 - f_t)^2/f_t]$.

demonstration alone. The $\chi^2$ and corresponding $P$ values provide the investigator with sufficient material to draw conclusions about the differential effects of the two methods of instruction under the varied conditions of motivation, recall, and retention, puzzle content, and type. Since the purpose here is to describe a method of analysis and not to draw conclusions from hypothetical data, the reader will be spared the discussion outlined. However, it should be noted in passing that the results of the analysis appear in a concise, understandable form.

Table 3 presents an analysis of the same data in terms of the reliability of mean differences employing the conventional $t$ technique. Table 3 strongly substantiates Table 2 with the greatest difference being the somewhat higher confidence levels reflected in the $P$ values of Table 3. A comparison in terms of the usual interpretation of $P$ values (.01 highly significant, .05 significant, > .05 not significant), indicates that in six cases out of twelve the interpretation would be the same, in five cases there would be a one-step difference, and in one case a two-step difference. The similarity of the two methods is further indicated by the rank correlation coefficient between

TABLE 3

Analysis in Terms of Mean Differences between Two Methods of Instruction
(Based on hypothetical data from Table 1)

| Experimental condition | $M_1$ Dem | $M_2$ Dem Exp | $M_d$ | $\sigma_d$ | $\sigma_{M_d}$ | $t$* | $df$ | $P$ |
|---|---|---|---|---|---|---|---|---|
| Motivation | | | | | | | | |
| Reward | 29.10 | 26.91 | 2.19 | 2.45 | .48 | 4.56 | 26 | <.01 |
| No reward | 29.76 | 28.75 | 1.01 | 1.77 | .35 | 2.89 | 26 | <.01 |
| Test | | | | | | | | |
| Day 1 | 31.53 | 30.27 | 1.26 | 2.35 | .57 | 2.21 | 17 | .05—.02 |
| Day 2 | 28.42 | 26.34 | 2.08 | 2.33 | .57 | 3.65 | 17 | <.01 |
| Day 3 | 28.32 | 26.88 | 1.44 | 1.86 | .45 | 3.20 | 17 | <.01 |
| Puzzle content | | | | | | | | |
| Numbers | 25.51 | 25.36 | .15 | 1.59 | .39 | .38 | 17 | .80—.70 |
| Patterns | 28.82 | 26.27 | 2.55 | 2.38 | .58 | 4.40 | 17 | <.01 |
| Symbols | 33.94 | 31.87 | 2.07 | 1.83 | .44 | 4.70 | 17 | <.01 |
| Puzzle type | | | | | | | | |
| 1 step | 18.21 | 17.72 | .49 | .94 | .23 | 2.13 | 17 | .05—.02 |
| 2 step | 25.46 | 24.26 | 1.20 | 1.69 | .41 | 2.93 | 17 | <.01 |
| 3 step | 44.62 | 41.51 | 3.11 | 2.71 | .66 | 4.71 | 17 | <.01 |
| Total experiment | 29.43 | 27.83 | 1.60 | 2.22 | .30 | 5.33 | 53 | <.01 |

* $t = \frac{M_1 - M_2}{\sigma_d/\sqrt{N-1}}$.

chi-square and $t$ of .91. If the square of this value may be taken to indicate the amount of overlap in the two methods of analysis, then the frequency of difference analysis is approximately eighty per cent as effective as the $t$ test. It would be pertinent to note at this point that the hypothetical data were set up systematically and were not juggled in any way later to produce greater correspondence.

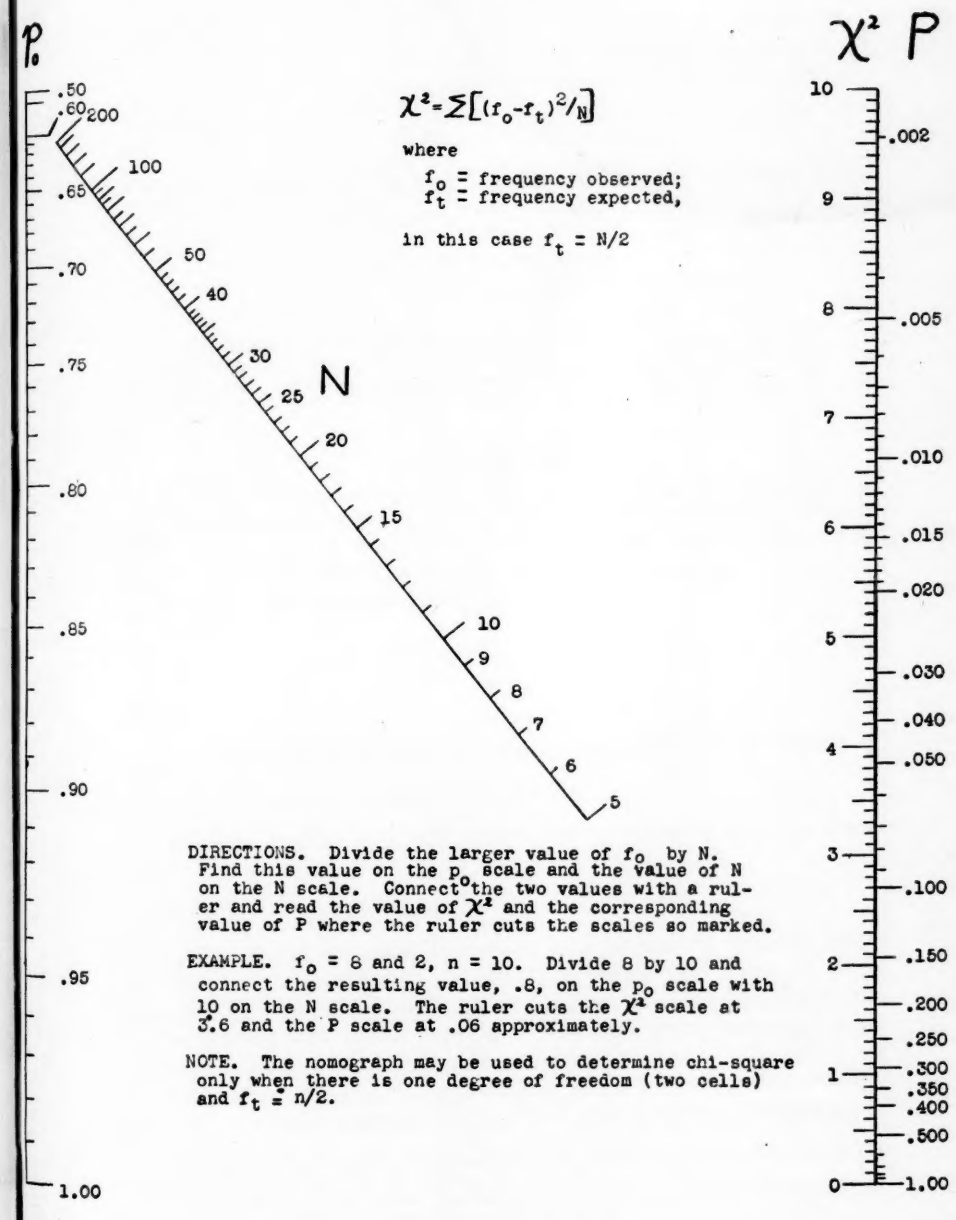
Whatever sacrifice in precision is entailed, the saving in computing time is considerable. The frequency of difference analysis reported here was done in less than thirty minutes without the aid of a calculating machine, whereas the mean difference analysis required about three hours with a calculator. As regards the alternate technique of analysis of variance, one thing is certain — it would have taken far more time. An even greater saving of time may be accomplished with the aid of the nomograph here provided. It is designed to determine the value of chi square when there is one degree of freedom and $f_t = n/2$. The nomograph includes a table of $P$ values corresponding to chi square.

The reader will readily visualize additional applications of the frequency of differences method. It is applicable wherever comparative

measures of effectiveness are available for methods, instruments, aids, chemical compounds, etc. As was noted previously, the measures of effectiveness may be scores on tests or other measuring instruments, time scores, or statistics based upon individual scores. Some experimental applications may take advantage of the fact that the technique takes into account only the frequency and direction of the differences and disregards the magnitude of the differences.

Whatever precision is lost is due to this disregard of magnitude. However, this disadvantage is offset by a number of advantages which may be listed as follows:

1. The method is extremely rapid, involving little computation and hence little opportunity for error.
2. It has a wide range of applications, particularly in the field of research on training methods and aids.
3. It yields results which are readily understandable and can be explained to those unfamiliar with statistical methods.
4. It involves no terminology or concepts outside the realm of conventional statistics.




$$\chi^2 = \Sigma\left[(f_o - f_t)^2/N\right]$$

where

$f_o$ = frequency observed;
$f_t$ = frequency expected,

in this case $f_t = N/2$

DIRECTIONS. Divide the larger value of $f_o$ by N. Find this value on the $p_o$ scale and the value of N on the N scale. Connect the two values with a ruler and read the value of $\chi^2$ and the corresponding value of P where the ruler cuts the scales so marked.

EXAMPLE. $f_o$ = 8 and 2, n = 10. Divide 8 by 10 and connect the resulting value, .8, on the $p_o$ scale with 10 on the N scale. The ruler cuts the $\chi^2$ scale at 3.6 and the P scale at .06 approximately.

NOTE. The nomograph may be used to determine chi-square only when there is one degree of freedom (two cells) and $f_t = n/2$.

PSYCHOMETRIKA—VOL. 12, NO. 1
MARCH, 1947

# AN INDEX OF ITEM VALIDITY PROVIDING A CORRECTION FOR CHANCE SUCCESS

A. P. JOHNSON
PURDUE UNIVERSITY

The KG Index described below is proposed for evaluation as one approach to the problem of providing an index giving comparable values for items (1) of equal discriminative power at all levels of difficulty and (2) of different numbers of alternative responses.

## 1. *The KG Index*

In 1934 Votaw* suggested that item validity comparisons for upper and lower 27% groups of a tested population sample be based on the proportion *in each group who know the answer* to an item rather than on the proportion who *answer it correctly.*

He gave the general equation:

$$x = \frac{nR - N}{n - 1},$$

in which

$x$† = the number in each group who *know* the correct answer,

$n$ = the number of choices in the item,

$R$ = the number of correct responses to the item within a given group,

$N$ = the number of cases *in each group.*

Votaw found in some instances that values of $x$ were negative. He construed these to mean that either the items in question were (1) so stated as to "trick" examinees into making incorrect responses, (2) keyed incorrectly, or (3) suffering from some other serious fault rendering them invalid.

Without presenting the generalized ratio, he shows that for any group all of whom respond to any item and *who are completely in*

* Votaw, D. F. Notes on validation of test items by comparison of widely spaced groups. *J. educ. Psychol.*, 1934, 25, 185-191.

† The writer uses hereafter the symbol $K$ for the number of the total test sample who are estimated to *know* the correct answer.

*ignorance of the correct answer* $N/n$ of them would be expected to mark the correct answer by chance. An experimental study by Ruggles of the marking by a student group of the correct choice in a true-false test the material of which was wholly unfamiliar to them is mentioned by Lee and Symonds.* The percentage of correct responses actually obtained was 51% when the chance expectation was 50%. Votaw shows that the probability of any invalidating conditions existing in an item can be determined for any given negative value of $x$ for that item by considering $N/n$ plus or minus its probable error. He reported the P.E. of $N/n$ to be equal to $.6745\sqrt{Npq}$, where $p = 1/n$ and $q = (1 - 1/n)$.

In 1936, Guilford† proposed as a means of evaluating the level of difficulty of a test item the proportion passing corrected by an allowance for chance success. This corrected proportion, $c_p = (nR - N)/N(n - 1)$, is simply the proportion of the total test population who may be expected to *know* the correct answer. In short, based on the *total* $R$ for the item it is $K$ (or Votaw's $x$ *for the entire test population*) divided by $N$.
Thus

$$K = \frac{nR - N}{n - 1}. \tag{1}$$

It is proposed that there be considered a new index, the KG Index, based not on contrasted upper and lower groups of 25%, 27%, or 33%, but on contrasted upper and lower groups equal in size *for each item to the number of individuals estimated to know the correct response to that item.* Suppose for a given item that

$$\begin{aligned} R &= 200, \\ W &= 300, \\ N &= 500, \\ n &= \ \ 5. \end{aligned}$$

Then

$$K = \frac{nR - N}{n - 1} = \frac{5 \times 200 - 500}{4} = \frac{500}{4} = 125.$$

It is postulated that if those 125 persons who are estimated to know the correct answer to the given item are the 125 persons who are *highest* on the criterion scale, that item may be said to have a *perfect*

* Lee, J. M. and Symonds, P. M. New type or objective tests: A summary of recent investigations (October 1931-1933). *J. educ. Psychol.*, 1934, 25, 161-184.

† Guilford, J. P. The determination of item difficulty when chance success is a factor. *Psychometrika*, 1936, 1, 259-264.

*positive* relationship to the criterion. If those 125 who were estimated to know the correct answer are the 125 persons who are *lowest* on the criterion scale, that item may be said to have a *perfect negative* relationship to the criterion. The extent of this relationship could be determined by arranging the 500 papers in decreasing order of criterion scores and determining for that given item how many correct responses occurred among the top 125 papers. The more nearly that item approached perfect positive relationship with the criterion the more nearly the number of correct responses among the top 125 papers would, in this instance, approach 125. If this item had a perfect negative relationship to the criterion scores, all 125 who were estimated to know the correct answer would appear among the lowest 125 on the criterion scale. The remaining number of correct answers, $R - K$, 200 — 125 or 75, would be distributed among the 500 — 125 or 375 papers remaining. Chance expectations with 75 rights among 375 papers would be 1 in 5 throughout the upper 375 papers on the criterion scale. Thus the upper 125 papers would be expected to include about 125/5 or 25 correct responses by chance.

The KG Index can be developed as follows as a convenient summary of how closely the actual responses approximate the perfect (or deviate from the chance) relationship with the criterion scores, as that relationship is defined above.

Let us use the symbols, $R_U$ for the number of right responses in the upper group and $R_L$ for the number of right responses in the lower group. In the *perfect positive* relationship postulated above the ratio $R_U/K$ should equal 1.0, and the ratio $R_L/K$ should equal $1/n$ (since $R_L$ should equal $K/n$).*

In a *chance* relationship both $R_U/K$ and $R_L/K$ should equal $1/n$.†

In a *perfect negative* relationship as postulated above, the ratio $R_U/K$ should equal $1/n$‡ (since $R_U$ should equal $K/n$), and the ratio $R_L/K$ should equal 1.0.

It is possible to obtain a positive index for so-called positive relationship, a zero index for chance relationship and a negative index for the so-called negative relationship by subtracting the ratios $R_U/K$ and $R_L/K$. This difference of proportions, $R_U - R_L/K$, is essentially the same as the U-L Index proposed by the writer. The two are iden-

* As Votaw indicates, the value $N/n$ or in this instance $K/n$ may well vary, as estimated by the formula S. E. of $K/n = \sqrt{K \times 1/n \times (n-1)/n}$. According to the probability tables for the normal curve of error, in 9973 cases out of 10000, $K/n$ should not vary beyond $\pm 3\sqrt{K \times 1/n \times (n-1)/n}$. It is thus possible that values of less than $1/n$ may occur.

† Idem.

‡ Idem.

tical when the value of $K$ is $.27N$. The difference between $R_U - R_L/K$ and the U-L Index is that $K$ is not fixed at $.27N$ but may vary from 0 to $N$. Whenever $K$ exceeds $N/2$, however, the upper and lower $K$ groups overlap by the amount $2(K - N/2)$ or $2K - N$. For the sake of simplicity, the practical question of how to handle omitted responses is deferred until later. If the group considered to be guessing the correct answer (i.e., those not knowing it) is designated by $G$, then $N - K = G$. If, when $K$ is greater than $N/2$, the $G$ group rather than the $K$ group is made the basis for upper vs. lower group comparison, the resulting difference between actual correct responses in upper and lower groups is the same and *the groups do not overlap.* The divisor $(G)$ then is no greater than the effective maximum value of the groups whose numbers of correct responses are subtracted. The symbol $B$ can be substituted for $K$ in the ratio $R_U - R_L/K$, where $B$ (i.e., the base group) is $K$, when $K$ does not exceed $N/2$, and $G$ when $K$ exceeds $N/2$.

Since for perfect positive relationship as defined above the expected value of $R_U - R_L/B$ is $1 - 1/n$ or $(n-1)/n$ and since for perfect negative relationship the expected value of $R_U - R_L/B$ is $1/n - 1$, the ratio in this form has a maximum theoretical value dependent on the number of choices. This dependence can be eliminated by multiplying the ratio $R_U - R_L/B$ by $n/(n-1)$. This expression is the KG Index:

$$\text{KG Index} = \frac{n(R_U - R_L)}{(n-1)B}. \tag{2}$$

### 2. *Computation of the KG Index*

As a first step, the test papers are arranged from highest to lowest on the criterion scale to be used as the standard of validity. The number of persons marking the correct response and the number marking all incorrect responses in the upper 30% of papers and in the lower 30% of papers is determined by graphic item count or other means. In order to provide most efficiently for data needed later, it is suggested that the papers be divided into successive groups as follows: upper and lower 6%, next upper and lower 4%, next upper and lower 5%, next upper and lower 5% and the remaining upper and lower 10% to total 30% in each.* With the graphic item counter, for instance, each upper group is run through separately and a separate count obtained on each. The same procedure is followed with the lower groups. For example, by adding the data of the 6% and the

* The use of these suggested groups is believed to provide sufficient accuracy while avoiding the necessity of computing different base groups for each item.

next 4% groups the information necessary for computing the KG Index based on the upper and lower 10% groups for certain specific items can be obtained. Similarly the data for any desired base groups from 6% to 30% can be found for each item.

The average number of correct and the average number of incorrect responses in the upper and lower 30% groups provide a practicable, if not a precise, means of determining the difficulty level of each item by the formula:

$$K = R - \frac{W}{(n-1)}. \tag{3}$$

Divided by .3 $N$ and when there are no omissions, this value becomes essentially equivalent to Guilford's $c_p$.

Table 1 gives a suggested range of values of $K$ in terms of $N$ for which upper and lower 30% groups provide convenient base groups and for which base groups of smaller size are suggested.

Table 1
Recommended Base Groups and Comparable Ranges of $K$ and $R$*

| Base Group | Range of $K$ in terms of $N$ | Range of $R$ in percentages of $N$ for specified numbers of choices | | | |
|---|---|---|---|---|---|
| | | $n=5$ choices | $n=4$ choices | $n=3$ choices | $n=2$ choices |
| .30$N$ | .25 — .75$N$ | 40.0 — 80.0 | 43.8 — 81.2 | 50.0 — 83.3 | 62.5 — 87.5 |
| | .18 — .24$N$ | 34.4 — 39.9 | 38.5 — 43.7 | 45.4 — 49.9 | 59.0 — 62.4 |
| .20$N$ | .76 — .82$N$ | 80.1 — 85.6 | 81.3 — 86.5 | 83.4 — 87.9 | 87.6 — 91.0 |
| | .13 — .17$N$ | 30.4 — 34.3 | 34.8 — 38.4 | 42.0 — 45.3 | 56.5 — 58.9 |
| .15$N$ | .83 — .87$N$ | 85.7 — 89.6 | 86.6 — 90.2 | 88.0 — 91.3 | 91.1 — 93.5 |
| | .08 — .12$N$ | 26.4 — 30.3 | 31.0 — 34.7 | 38.6 — 41.9 | 54.0 — 56.4 |
| .10$N$ | .88 — .92$N$ | 89.7 — 93.6 | 90.3 — 94.0 | 91.4 — 94.7 | 93.6 — 96.0 |
| | .04 — .07$N$ | 23.2 — 26.3 | 28.0 — 30.9 | 36.0 — 38.5 | 52.0 — 53.9 |
| .06$N$ | .93 — .96$N$ | 93.7 — 96.8 | 94.1 — 97.0 | 94.8 — 97.3 | 96.1 — 98.0 |
| | 0 — .03$N$ | 20.0 — 23.1 | 25.0 — 27.9 | 33.3 — 35.9 | 50.0 — 51.9 |
| 0 | .97 - 1.00$N$ | 96.9 — 100 | 97.1 — 100 | 97.4 — 100 | 98.1 — 100 |

(Underlined values represent expected percentages correct when *all answers* are marked according to chance.)

* The suggested base group values of this table have been arrived at on the basis of both theoretical and practical considerations; except for a very few special test construction situations it is expected that they will prove quite satisfactory.

The values of $R$ have been derived from the basic formula $K = (nR - N/(n-1)$ solved for $R$, namely, $R = (n-1)K/n + N/n$.

In most instances it is believed that the majority of items will be of such difficulty that the upper and lower 30% groups will serve. If the suggested breakdown of groups has been followed, the data necessary for computing the KG Index will be readily available for all items.

The following data for one item of a five-choice test will illustrate the method for computing the KG Index when the number of omissions is negligible:

$$\begin{aligned} R &= 268 \\ W &= 96 \\ 0 &= 4 \quad \text{(negligible)} \\ N &= 368 \\ n &= 5 \end{aligned}$$

$$K = R - \frac{W}{n-1} = 268 - \frac{96}{(5-1)} = 268 - 24 = 244.$$

In terms of $N$, $K = 244/368 = .663N$.

Entering the second column of Table 1, "Range of $K$ in terms of $N$," we note that $K$ for this item falls within the range $.25 - .75N$ corresponding to a base group of $.30\ N$ (see the first column).

Thus the desired base group equals $.30 \times 368$ or 110.4. This figure is rounded off to 110. From the total count for the upper 30% of the test papers and from the total count for the bottom 30% of the test papers, the number of correct responses in these groups is found. The actual counts were:

$$\begin{aligned} R_U &= 107 \qquad \text{base group, } B = 110 \\ R_L &= 47 \end{aligned}$$

$$\text{KG Index} = \frac{n(R_U - R_L)}{(n-1)B}.$$

Substituting, we have

$$\text{KG Index} = \frac{5(107-47)}{4 \times 110} = .68.$$

For this item the tetrachoric $r$ based on a 66% vs. 34% split was .67; the tetrachoric $r$ based on a 50% vs. 50% split was .60 and Guilford's $\phi^*$ based on contrasted top and bottom 25% groups was .66.

The agreement of these values is not always so close, as is illustrated by the data for a two-choice item where chance successes tend to attenuate the $\phi$ coefficient (and similarly other indices of relation-

* Guilford, J. P. The phi coefficient and chi square as indices of item validity. *Psychometrika*, 1941, 6, 11-19.

ship not making a correction for chance successes). For a specific two-choice item, $N = 401$, $R = 239$, $W = 158$, $0 = 4$, and since $K = .202N$, $B = .20N$. $R_U$ and $R_L$ were 62 and 36, respectively; thus $n(R_U - R_L)/(n-1)B = .65$. The corresponding $\phi$ was 1.26; no values of the tetrachoric $r$ were computed.

Formulas (1) and (2) are not applicable where the number of omissions is appreciable, for they assume no omissions.

Formula (3), $K = R - W/(n-1)$, can be used regardless of the number of omissions. It is useful to determine $K$ when it is necessary to item-analyze speeded tests in which the number of omissions among later items is appreciable. When $K$ is expressed in terms of $N$, $N$ should include only those who read the item, not the "non-reads" as defined below.

Omissions on speeded tests are usually of two types: (a) those by persons who read the item but fail to mark any choice, and (b) those by persons who have not read as far in the test as the end of that item. Frederick B. Davis in a communication to the writer suggests a method of determining the "non-reads" directly. The test papers are scored and arranged in rank order according to the criterion score. The top 6%, next 4%, next 5%, next 5%, and remaining 10%, to total the top 30% of the test papers, are segregated as are similar bottom groups. Next the papers of each upper group are gone through separately to determine which item on each paper is the last one marked. A tally is then made opposite the number of the next item, for that one is presumably *the first item* not read by the subject. This tally is the basis for a *cumulative* frequency table of the *"non-reads"* for each particular item among the top 6%, top 10% (6% + 4%), top 15%, top 20%, and top 30% of the test papers. A similar count is made of the "non-reads" among the lower 6%, 10%, 15% 20%, and 30% groups.

By graphic item count on the International Test Scoring machine or by other means, the number of right responses and the number of wrong responses in each top group to each item is found. The omits may be obtained, if desired, by subtracting from the number of papers within the appropriate upper group the rights plus wrongs plus "non-reads." Similarly the omits among the several lower groups may be obtained, if desired.

Where the "non-reads" represent an appreciable proportion of the appropriate base groups, the following modification of the basic formula for the KG Index is suggested:

$$\text{KG Index} = \left(\frac{n}{n-1}\right)\left[\frac{R_U}{(B - \text{"non-reads"}_U)} - \frac{R_L}{(B - \text{"non-reads"}_L)}\right], \quad (4)$$

in which $n$, $R_U$, $R_L$, and $B$ have the same meanings as in Formula (2),

"non-reads"$_U$ = "non-reads" in upper base group, and
"non-reads"$_L$ = "non-reads" in lower base group.

The postulated standard of perfect positive and of perfect negative relationship against which each item is evaluated by the KG Index is based upon probability theory. In the strictest sense, it is valid only when (a) the number of cases is large and (b) when all of the alternative responses are equally enticing to those examinees in ignorance of the subject matter of the question. Although these conditions are not too frequently met in practice, the KG Index is believed to possess some possible usefulness as an item validity index giving closely comparable values for items (1) of different levels of difficulty but equal discriminative power and/or (2) of different numbers of alternative responses.

PSYCHOMETRIKA—VOL. 12, NO. 1
MARCH, 1947

## BOOK REVIEWS

HAROLD CRAMER *Mathematical Methods of Statistics.* Princeton: Princeton University Press, 1946. Pp. xvi + 575.

It is the purpose of the author to present a logical development of the method of mathematical statistics, which presupposes on the part of the reader a mathematical knowledge of only calculus, algebra, and analytic geometry. The first section of the book, containing 137 pages, is devoted to the presentation of the various topics in higher mathematics which are necessary for the proof of theorems in the later sections of the book. This section, beginning with point-set theory, contains a logical development of Lesbegue measure, the Lesbegue integral, theory of additive set functions, and the Lesbegue-Stieltjes integral. The procedure throughout is to make a detailed, rigorous development for the simplest case and to indicate briefly the possible generalizations of the theorems to less restricted conditions and to multidimensional space. At the end of the first part of the book, one chapter each is devoted to characteristic functions and matrix theory. A final chapter of this part covers such topics as Stirling's formula and Beta and Gamma functions.

This first section of the book may be understandable to European students who have finished a course in calculus, but in terms of American education in mathematics, it requires more background than that. Especially it requires a familiarity with the rigorous development of the calculus and with the manipulation of the functions of complex variables. The first section is, nevertheless, extremely valuable because the selection of pertinent material from function theory provides a background for statistics which would be very difficult for anyone who is not a professional mathematician to obtain.

The second part of the book discusses random variables and probability functions, both univariate and multivariate. The two introductory chapters deal with the fundamental basis of probability theory. Cramér develops the concept of probability from the empirical knowledge of frequency ratios but does not actually base his mathematics upon frequency ratios. Instead probability is a mathematical model, a function of point sets, consistent within itself, but designed to have a reasonable correspondence with the properties of frequency ratios. The discussion of random variables includes the important binomial, Poisson, normal, chi-square, $t$, $z$, and incomplete Beta distributions as well as explanations of Gram-Charlier series and the Pearson system of distributions. General properties of multivariate distributions are also discussed.

The third section, on statistical inference, includes two main types of material, general theory of statistical tests and inferences and the details of particular sampling distributions. The section on the theory of estimation and the chapter on the theory of testing hypotheses are particularly valuable. The discussion of the topic, however, seems less complete than the other sections of the book. Cramér probably felt that the other material was mathematically more fundamental.

For psychologists the book will be very useful, almost indispensable, to those who are trying to develop real mathematical sophistication in statistics. After

careful study of Cramér, the reader should be in a position to understand more advanced texts and the periodical literature. For the person who wishes to become a sophisticated user of statistics, but not necessarily a mathematical statistician, Cramér is also valuable but less so. His discussion is neither complete enough nor non-technical enough to give such a reader an understanding of the full implications of such concepts as "uniformly most powerful test," or "unbiassed test." The book is not, nor was it intended to be, a handbook of statistical tests. In that sense also it is not of maximum value for the practical statistician. There is, however, such a scarcity of integrated discussions of the modern theory of statistical tests and inferences that even for the less mathematical reader the book is valuable.

*Fels Research Institute for the Study of Human Development*
ALFRED L. BALDWIN

DAVIS, FREDERICK B. *Item-Analysis Data: Their Computation, Interpretation, and Use in Test Construction.* Harvard Education Papers Number 2. Cambridge: Graduate School of Education, Harvard University, 1946. Pp. v + 42.

This monograph does not undertake to provide a complete review and discussion of techniques of item-analysis. It is limited to (1) the exposition of one procedure which the author has developed for analyzing and expressing the difficulty and discriminating power of items and (2) critical remarks on certain specific problems in the interpretation and use of item-analysis data. In addition, a four-page bibliography on item-analysis is provided.

The distinctive feature of the procedures presented by the author is that they yield indices scaled in presumably equal units, with a range from 0 to 100. In the case of difficulty, defined as per cent of the group knowing the answer to an item, this involves assuming that ability is normally distributed in the group studied and then converting percentages into abscissa values of the normal probability curve. These are then multiplied by an appropriate constant to give the desired range of scores. In the case of discrimination indices, correlation coefficients are translated into values of Fisher's $z$, and these are multiplied by the constant which yields a value of 100 corresponding to 99 per cent success in the upper 27 per cent of cases and 1 per cent success in the lower 27 per cent.

Both difficulty and discrimination indices are extracted from the per cent of successes, failures, and omissions in the top and bottom 27 per cent of the group on the criterion measure (usually total score). A chart has been prepared which provides the difficulty and discrimination indices for any pair of percentages of success (after correction for chance) in the two groups. Thus, the procedure and the table are an elaboration of those developed by Flanagan earlier.

The procedure of obtaining item indices from per cent of success in the upper and lower 27 per cent has been found to be an efficient and practical procedure, especially where an IBM test-scoring machine with a graphic item counter attachment is available. The use of scaled values for difficulty and discrimination indices gets away from the non-linearity of the number scale both of proportions and of correlation coefficients. However, the numeral values of the proposed new indices will be entirely unfamiliar to the user and will present some difficulty of interpretation on that account. Their value would appear to be chiefly for individuals who are going to do a great deal of item analysis and in cases where the item indices

are to be used for comparative purposes within a test-development organization. The indices would probably prove a source of confusion in published reports.

Chapter IV of the monograph presents a stimulating discussion of various problems connected with the use and interpretation of item-analysis data. In general, the tenor of these remarks is to emphasize that item-analysis is a valuable supplementary aid to but not a substitute for good item-writing and editing, and that item-analysis data should be used with insight and discretion rather than mechanically. Though the discussions are brief and suggestive rather than definitive in many cases, they should stimulate the reader to critical thought on a number of phases of the item-analysis problem.

*Teachers College, Columbia University* ROBERT L. THORNDIKE

## SPECIAL NOTICE

The U. S. Civil Service Commission has announced an examination for filling Research Psychologist positions in Washington, D. C., and throughout the United States.

The salaries for Research Psychologist positions range from $4,902 to $9,975 a year. The duties of the positions are of a highly responsible and technical nature. To qualify, applicants must have had 4 years of progressive professional experience in conducting or participating in important research projects in the field of psychology. This experience must have been at successively higher levels of responsibility for the higher grades, and must show the applicant's ability to plan, direct and coordinate research programs of considerable scope and complexity. Applicants for the highest salary level must have earned recognition as leaders in the field of psychology. Graduate study in psychology may be substituted year for year for 3 years of the required experience. No written test is required for this examination. The age limit of 62 years is waived for persons entitled to veteran preference.

Applications for the Research Psychologist examination will be accepted until further notice. However, some positions will be filled immediately. Persons interested in these positions should apply at once. Information and application forms may be obtained at most first- and second-class post offices, from Civil Service regional offices, and from the U. S. Civil Service Commission, Washington, D. C.